Optimisation of rate-pitch perception in cochlear implant hearing

Andrew E. Vandali

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

2013

Department of Audiology and Speech Pathology
Faculty of Medicine, Dentistry and Health Sciences
The University of Melbourne

Produced on archival quality paper
Abstract

It is well established that Cochlear implants (CIs) enable moderate-to-profoundly deaf people to understand speech without the aid of lip reading. However, the perception of voice and musical pitch by implantees is far from satisfactory. The limitations stem from an inability to code and perceive detailed information about the fundamental frequency (F0) in signals, which in normal hearing provides the principle cue to pitch. Present CI systems instead focus on conveying information about the signal’s envelope. In these systems, F0-pitch (albeit somewhat weaker than that of normal hearing) is coded over a limited range by way of amplitude modulation in the temporal (time) envelope. However, the depth and shape of this modulation is highly dependent on acoustic properties of the signal and characteristics of the sound coding strategy. Thus the salience and accuracy of coded F0 information can vary substantially across signals, acoustic environments, and coding techniques. Furthermore, variations in spectral (frequency) envelope information (e.g., differences amongst vowels) which produce changes in the place of electrical stimulation, can adversely affect CI recipients’ judgement of pitch derived from temporal envelope information. The main aims of the research were: (1) to develop a rate-pitch coding strategy that enhances the salience (pitch-strength) and accuracy (pitch-height) of coded F0 temporal information without adversely affecting recognition of speech; and (2) to explore novel methods in which individual’s ability to attend to F0 information exclusively as a cue to pitch in the presence of spectral envelope variations are improved through the use of a training program.

To address the first aim, psychophysics experiments were conducted with CI recipients using modulated electrical pulse trains to examine the effect of modulation depth, rate, and shape on pitch-height and loudness relative to that of unmodulated electrical pulse trains. Unmodulated pulse trains were used as a reference because previous studies had shown that accurate identification of pitch intervals could be obtained on the basis of changes in pulse rate, at least for low rates. Results of the present psychophysics experiments demonstrated that deep sinusoidal amplitude modulation was needed to elicit an accurate percept of pitch for low F0 signals. For a sharper modulation function, similar performance could be obtained at a shallower modulation depth. However, application of deep or sharp modulation reduced loudness. These results, together with outcomes of previous studies, provided input necessary for
the development of an experimental strategy (eTone) that enhanced coding of F0 modulation in the stimulus envelope. Pitch and speech perception using that experimental strategy were compared to that of the clinical Advanced Combinational Encoder (ACE) strategy in six adult CI recipients. Significant improvements in discrimination of pitch were observed with no reductions in recognition of speech in quiet or noisy conditions.

The second aim of the research was to determine whether improvements in pitch perception could be obtained through specific training that primarily directed listeners to attend to F0 information exclusively as a cue to pitch, and to resonant frequency as a cue to spectral timbre. Outcomes demonstrated significant improvements in CI recipient’s F0 discrimination thresholds after training whereas no improvements were observed for control subjects that did not participate in training. This outcome generalised across different stimuli including natural sung vowels that embodied small variations in spectral timbre that could adversely affect judgement of F0 pitch. The improvement was also maintained several months after completion of training.
Declaration

This is to certify that:

(i) the thesis comprises only my original work towards the PhD except where indicated in the Preface,
(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, exclusive of tables, maps, bibliographies and appendices.

Name: Andrew E. Vandali

Signed:………………………………….   Date:……………………..
Preface

This research was supported by the HEARing CRC, established under the Cooperative Research Centres Program – an Australian Government Initiative. In addition, we acknowledge the support of the Bionics Institute and the support it receives from the Victorian Government through its Operational Infrastructure Support Program.

Ethical approval for the research was obtained through the Human Research Ethics Committee of the Royal Victorian Eye and Ear Hospital (RVEEH). Clinical management and recruitment of cochlear implant recipients was provided by the RVEEH Cochlear Implant Clinic.

Four journal publications were generated through the research along with a number of conference presentations (see list of publications). Co-authors on those publications and presentations comprised the supervisors of the present doctoral study. Chapters 3-5 within the thesis were derived almost verbatim from three of those journal articles.

Some development of the experimental strategy evaluated in the present research was conducted prior to commencement of the doctoral study (see patent applications in list of publications).

Subject recruitment and data collection for the pitch training study in normal-hearing listeners reported in the present research was conducted under supervision of the present PhD candidate by a Masters of Clinical Audiology student at the University of Melbourne (see list of publications).

Financial support for travel and accommodation at the 2011 Conference on Implantable Auditory Prosthesis was provided by the sponsors of that meeting who included Advanced Bionics, Med-El, the National Institutes of Health, Cochlear Corporation, House Ear Institute, Deafness Research UK, University College London, and the University of Wisconsin.
List of Publications

Peer-reviewed journal publications


Conference presentations


**Patent applications**


**Masters of Clinical Audiology project – Minor thesis**

Acknowledgements

When I first began working in this field some twenty years ago my research activities were not hindered by the fact that I lacked a PhD qualification. However as time passed, difficulties related to securing of independent funding and management of students as a principle supervisor became increasingly problematic. About six years ago a project came to mind that seemed very appropriate for a doctoral study. After sitting on the idea for a few years I decided to tackle the project myself. Along the way I gained new skills as a researcher, broached different fields of science, and learnt a lot about myself. Overall I found the experience very rewarding and I am thankful to all my colleagues, family members, and friends who helped to inspire and encourage me.

An instrumental source of motivation for this research came from all those wonderful cochlear implant recipients that I have had the pleasure to meet and work with over the years. These individuals donate time to participate in research because of their desire to give something back to the research bodies responsible for partial restoration of their hearing. Through working with these individuals, I became interested in trying to improve aspects of their hearing, beginning with perception of speech and more recently perception of pitch and its subsequent effect on their appreciation of music. So to all those recipients who have donated their time to this research, thank-you. Through your generous efforts to help us, we are motivated to help you.

Motivation is important, but just as valuable is having good supervisors. I was very fortunate to have three supervisors who each helped to shape the directions and outcomes of the research in their own way. First and foremost, the expertise and guidance provided by Richard van Hoesel was invaluable. We have worked together for more than twenty years and I have always valued his wisdom and judgement. His patience, attention to detail, ability to indentify problems, comprehend difficult topics, and most importantly make sense of my writing, which can often be rather long winded, all had significant impacts on the quality of this work. Besides all that he has been a good and understanding friend who has helped me greatly along the way. Secondly, David Sly brought a completely different set of skills to the project. His background in neurophysiology contributed to my knowledge of neural mechanisms underlying pitch perception. He helped me to focus on the bigger picture when I was stuck down in the fine detail and his comments regarding my writing and presentation techniques were
very useful and greatly appreciated. On top of that his knowledge of university procedures/requirements for PhD students was excellent which greatly facilitated the whole process. Finally, to Robert Cowan who accepted me as a student after I had spent more than a decade working under his direction as researcher engineer within the HEARing CRC. At the time I was finding it difficult to find a supervisor and he was prepared to take on that arduous job. He helped in terms of recruiting subjects, adhering to correct clinical and ethical practices, in writing of the thesis, and in general management of the research activities and infrastructure. Thank-you Richard, David, and Bob for your generous contributions to this research.

I would also like to thank Richard Dowel and James Fallon for their contributions as members of the advisory panel of this doctoral study. Both helped to steer research directions towards useful outcomes that could be achieved within the time frame of the study.

A lot of the research would not have been possible had it not been for the engineers that developed the SPEAR3 research processor and associated software tools. Those hardworking people are Mark Harrison, Richard van Hoesel, Justin Zakis, Hugh McDermott, and Rodney Millard. Many thanks also to Hugh McDermott and Justin Zakis for providing source code for a prototype implementation of the F0 estimator which was subsequently modified for the eTone strategy.

I’d also like to thank Ray Goldsworthy, Stephanie Wong, Aswin Wijetillake, and my children, Claire and Lucas, for their help with initial testing of the music-pitch training program. Thanks also to David MacFarlane, Colette McKay, and Valerie Looi for producing and editing some of the sung vowel stimuli used in the research and to Elise McDonald for her work with the music-pitch training program conducted as part of her minor thesis project.

There are many colleagues that I have had the pleasure to work with over the years and during the course of this doctoral study which I’d also like to thank. They include Mark White, Ian Bruce, Leon Heffer, and Laurie Cohen for input regarding neural pitch mechanisms. Many Thanks also to Ray and Radha Goldsworthy and Ed Ariel Hight for the many hours we have spent together sharing ideas and questioning current views related to cochlear implant hearing. Thanks also to Gregor Kennedy and Damien Smith for their comments regarding training paradigms, to Pam Dawson, Komal Arora, John Heasman, and Adam Hersbach for their help with implementation and testing of the latest version of the eTone strategy, to Brett Swanson, Harvey Dillon, and Jim Patrick,
for their general comments on the research, and to the organisers and steering committee of the 2011 Conference on Implantable Auditory Prosthesis for inviting me to present some of my research. I am also very appreciative of the ongoing support provided by the Robert Shepherd and his team at the Bionics Institute Australia.

None of this work would have been possible had it not been for the devoted support given to me by my family. They unquestioningly gave me with the time and space that I needed, showed great patience and understanding of my needs, particularly when my mind was elsewhere, and through their actions helped me to stay motivated and to complete this work in a timely fashion. You have my love always and ever.
# List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2AFC</td>
<td>two alternative forced choice</td>
</tr>
<tr>
<td>4AFC</td>
<td>four alternative forced choice</td>
</tr>
<tr>
<td>2I</td>
<td>two-interval</td>
</tr>
<tr>
<td>ACE</td>
<td>Advanced Combinational Encoder</td>
</tr>
<tr>
<td>ACF</td>
<td>auto-correlation function</td>
</tr>
<tr>
<td>ADRO</td>
<td>Adaptive Dynamic Range Optimisation</td>
</tr>
<tr>
<td>AGC</td>
<td>automatic gain control</td>
</tr>
<tr>
<td>ASC</td>
<td>automatic sensitivity control</td>
</tr>
<tr>
<td>ANOVA</td>
<td>analysis of variance</td>
</tr>
<tr>
<td>BPF</td>
<td>band-pass filter</td>
</tr>
<tr>
<td>BTE</td>
<td>behind-the-ear</td>
</tr>
<tr>
<td>CF</td>
<td>centre frequency</td>
</tr>
<tr>
<td>CI</td>
<td>cochlear implant</td>
</tr>
<tr>
<td>CIS</td>
<td>Continuous Interleaved Sampling</td>
</tr>
<tr>
<td>CNC</td>
<td>Consonant Nucleus Consonant</td>
</tr>
<tr>
<td>CU</td>
<td>clinical current unit</td>
</tr>
<tr>
<td>DL</td>
<td>difference limen</td>
</tr>
<tr>
<td>DR</td>
<td>dynamic range</td>
</tr>
<tr>
<td>DSP</td>
<td>digital signal processor</td>
</tr>
<tr>
<td>EDM</td>
<td>exponential decay modulation</td>
</tr>
<tr>
<td>EDR</td>
<td>electrical dynamic range</td>
</tr>
<tr>
<td>EL-level</td>
<td>equivalent-loudness level</td>
</tr>
<tr>
<td>EP-rate</td>
<td>equivalent-pitch rate</td>
</tr>
<tr>
<td>eTone</td>
<td>Enhanced-Envelope-Encoded Tone</td>
</tr>
<tr>
<td>F0</td>
<td>fundamental frequency</td>
</tr>
<tr>
<td>F1</td>
<td>first formant</td>
</tr>
<tr>
<td>F2</td>
<td>second formant</td>
</tr>
<tr>
<td>F3</td>
<td>third formant</td>
</tr>
<tr>
<td>F0mod</td>
<td>F0 modulation</td>
</tr>
<tr>
<td>F0Sync</td>
<td>F0 synchronised</td>
</tr>
<tr>
<td>FDL</td>
<td>frequency difference limen</td>
</tr>
<tr>
<td>FFT</td>
<td>Fast Fourier Transform</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>GLM</td>
<td>general linear model</td>
</tr>
<tr>
<td>HA</td>
<td>hearing aid</td>
</tr>
<tr>
<td>HPF</td>
<td>high-pass filter</td>
</tr>
<tr>
<td>LPF</td>
<td>low-pass filter</td>
</tr>
<tr>
<td>LSD</td>
<td>least significant difference</td>
</tr>
<tr>
<td>MD</td>
<td>modulation depth</td>
</tr>
<tr>
<td>MDE</td>
<td>Modulation Depth Enhancement</td>
</tr>
<tr>
<td>MEM</td>
<td>Multi-channel Envelope Modulation</td>
</tr>
<tr>
<td>MIPS</td>
<td>millions of instructions per second</td>
</tr>
<tr>
<td>MP</td>
<td>monopolar</td>
</tr>
<tr>
<td>MR</td>
<td>modulation rate</td>
</tr>
<tr>
<td>MS</td>
<td>modulation shape</td>
</tr>
<tr>
<td>MSP</td>
<td>Multi-peak</td>
</tr>
<tr>
<td>NH</td>
<td>normal hearing</td>
</tr>
<tr>
<td>PDT</td>
<td>Peak Derived Timing</td>
</tr>
<tr>
<td>PPS</td>
<td>pulses per second</td>
</tr>
<tr>
<td>PPS/CH</td>
<td>pulses per second per channel</td>
</tr>
<tr>
<td>RF</td>
<td>radio frequency</td>
</tr>
<tr>
<td>RMS</td>
<td>root mean square</td>
</tr>
<tr>
<td>SAM</td>
<td>sinusoidal amplitude modulation</td>
</tr>
<tr>
<td>SMSP</td>
<td>Spectral Maxima Sound Processor</td>
</tr>
<tr>
<td>SNR</td>
<td>signal-to-noise ratio</td>
</tr>
<tr>
<td>SPEAK</td>
<td>Spectral Peak</td>
</tr>
<tr>
<td>STAR</td>
<td>Spike-based Temporal Auditory Representation</td>
</tr>
</tbody>
</table>
List of Symbols and Terms

General symbols

§ section
d’ d-prime (dissimilarity index)

eTone symbols

Channel signals

n channel number

N_{ch} number of channels

E[n] channel envelope signals

E_S[n] smoothed (slow-varying) channel envelope signals

E_M[n] modulated channel envelope signals

E_{WB}[n] wide-bandwidth channel envelope signals

FFT phase vocoder

k analysis frame rate

t analysis frame number

b FFT bin number

p_{0}[b] average FFT bin powers

f_{k}[b] average FFT bin frequencies

P_{F0}[b] average FFT bin powers for current F0 estimate

F_{F0}[b] average FFT bin frequencies for current F0 estimate

MaxF maximum bin frequency

F0 estimator

Stage one of five processing blocks in the F0 estimator

c candidate F0 sieve number

i harmonic number

h harmonic number for estimated F0

F{0}_{EST} estimated fundamental frequency

STR harmonic signal-to-total power ratio

E_{F0}[c] estimated F0 for candidate F0 sieves

M_{F}[c] matched power
$S_P[c]$ estimate of harmonic signal power
$N_P[c]$ estimate of noise power
$T_P$ total power
$S_{BW}[c]$ summed bandwidth of sieve filters
$T_{BW}$ total bandwidth
$K_{BW}$ signal-noise bandwidth compensation factor
$W_{S_P}[c]$ weighted harmonic signal power
$W_{STR}[c]$ weighted signal-to-total power ratio
$W[c]$ weighting function
$K_W$ weighting function factor
$G$ Gaussian function
$K_G$ standard deviation constant for Gaussian function
$BW_{Max}$ maximum bandwidth of Gaussian function

Harmonic probability estimator
$STR_{ch}[n]$ harmonic signal-to-total power ratio
$HP_{ch}[n]$ harmonic probability
$F_{ch}$ frequency range of channel
$P_{ch}(f)$ estimate of harmonic signal power within FFT frequency bins

Channel modulation
$M$ modulation function
$K_{HP}$ modulation gain constant

aTune terms
Stage a training method defined by cue(s) trained and stimulus type
Level training variants within each Stage that successively increase in complexity/difficulty

Trial one repeat of the pattern matching procedure
Run block of Trials to assess performance within a Level
Token a single auditory or visual representation a tone
Pattern  
a sequence of Tokens of predefined length (equal to the number of tones in the sequence to be matched)

Block  
one or more connected visual Patterns

NumF0s  
number of different F0s within a Level

NumTimbres  
number of different timbres within a Level

NumTones  
number of different tones presented within a Level

NumSequentialTones  
number of sequentially presented Tokens within a Pattern

NumCombinations  
number of combinations of tones within a Pattern

ResponseTime  
estimated time to respond to presentation of acoustic Pattern

PresentationTime  
time required to present acoustic Pattern

ISI  
inter-stimulus interval or inter-tone interval

PatternsPerBlock  
number of Patterns within a Block

NewBlockDelay  
time delay between additions of a new Block

TokenVelocityScaler  
scalar to control velocity of visual Tokens

RewardThreshold  
number of consecutively correct responses required to obtain a reward

FastResponseTime  
maximum response time before fast response level decays

Score  
total score

NumCorrect  
number of responses correct excluding those that are unambiguous

DoubleScoreNumCorrect  
same as NumCorrect but when in double score mode

Bonuses  
score bonus multiplier

DoubleScore  
double score multiplier

Interval  
F0 or resonant frequency interval between stimuli
# Contents

Abstract ......................................................................................................................... i
Declaration ...................................................................................................................... iii
Preface .............................................................................................................................. iv
List of Publications .......................................................................................................... v
Acknowledgements ......................................................................................................... vii
List of Abbreviations ....................................................................................................... x
List of Symbols and Terms ............................................................................................... xii

## 1 Introduction ............................................................................................................. 1
  1.1 Overview .................................................................................................................. 1
  1.2 Statement of research questions ............................................................................. 5
    1.2.1 Aims .................................................................................................................. 5
    1.2.2 Hypotheses ...................................................................................................... 5
  1.3 Structure of thesis (chapter outline) ...................................................................... 6

## 2 Background and Relevance to Research ................................................................ 8
  2.1 Cochlear implants and sound coding strategies ...................................................... 8
  2.2 Properties of musical and voiced tones .................................................................. 10
    2.2.1 Pitch ............................................................................................................... 11
    2.2.2 Timbre .......................................................................................................... 11
  2.3 Pitch coding in cochlear implants ......................................................................... 12
    2.3.1 Place-pitch .................................................................................................... 16
    2.3.2 Rate-pitch .................................................................................................... 19
    2.3.3 Amplitude modulation pitch ....................................................................... 26
    2.3.4 Relevance to research .................................................................................. 31
  2.4 Music perception using clinical processing strategies ......................................... 32
    2.4.1 Pitch perception ............................................................................................ 32
    2.4.2 Melody recognition ....................................................................................... 36
    2.4.3 Timbre perception ....................................................................................... 38
    2.4.4 Summary ........................................................................................................ 40
  2.5 Tonal language perception using clinical systems ................................................. 41
    2.5.1 Mandarin ....................................................................................................... 42
    2.5.2 Cantonese ..................................................................................................... 45
    2.5.3 Effects of stimulation rate .......................................................................... 46
    2.5.4 Summary ........................................................................................................ 46
  2.6 Pitch coding strategies ............................................................................................ 47
    2.6.1 Place-pitch coding strategies ........................................................................ 47
    2.6.2 Rate-pitch coding strategies ......................................................................... 50
    2.6.3 Amplitude modulation rate-pitch coding strategies ..................................... 51
    2.6.4 Fine temporal structure coding strategies .................................................. 57
    2.6.5 Relevance to Research ............................................................................... 58
  2.7 Training to improve perception in cochlear implant hearing ......................... 60
2.7.1 Previous experience .................................................. 60
2.7.2 Auditory training in normal-hearing listeners .......... 60
2.7.3 Auditory training in cochlear implant recipients ......... 64
2.7.4 Relevance to research .......................................... 67

3 Pitch and Loudness of Amplitude-Modulated Stimuli ......... 69
3.1 Introduction .................................................................. 70
3.2 General Methods ...................................................... 75
  3.2.1 Subjects ............................................................... 75
  3.2.2 Electrical stimuli .................................................. 75
  3.2.3 Acoustic stimuli .................................................... 76
  3.2.4 Threshold and comfortable loudness levels .......... 76
  3.2.5 Loudness balancing ............................................. 77
  3.2.6 Data analysis ....................................................... 78
3.3 Exp. 1: Pitch of unmodulated electrical stimuli ............... 78
  3.3.1 Exp. 1a: Pulse rate difference limens ................... 78
  3.3.2 Exp. 1b: Pitch comparison as a function of electrode place differences 80
  3.3.3 Exp. 1c: Pitch comparison as a function of differences in stimulation level ........................................83
3.4 Exp. 2: Pitch and loudness of modulated electrical stimuli .... 85
  3.4.1 Exp. 2a: Effects of modulation rate, depth and shape on pitch and loudness ................................. 85
  3.4.2 Exp. 2b: Effects of presentation level on pitch and loudness ............................................. 92
  3.4.3 Analysis of effects of stimulation level differences on pitch matching ........................................ 95
3.5 Exp. 3: Pitch of modulated acoustic and unmodulated electrical stimuli .............................................. 96
  3.5.1 Exp. 3a: Effects of acoustic modulation rate .............. 97
  3.5.2 Exp. 3b: Effects of electrical place of stimulation .......... 99
3.6 General Discussion ................................................... 100
  3.6.1 General outcomes ............................................... 100
  3.6.2 Relevance to sound coding strategies .................... 101
  3.6.3 A conceptual neural mechanism explaining temporal pitch ................................................. 104
  3.6.4 Differences between subjects ............................... 106
3.7 Conclusion .................................................................. 107

4 Development of a Rate-Pitch Strategy (eTone) .................. 109
4.1 Introduction ............................................................. 109
4.2 Methods ................................................................. 112
  4.2.1 The eTone strategy ............................................ 112
  4.2.2 F0 estimator analysis .......................................... 126
  4.2.3 STR and harmonic probability analysis ................. 127
4.3 Results ..................................................................... 128
  4.3.1 Electrical stimulus patterns ................................. 128
  4.3.2 F0 estimator analysis .......................................... 130
  4.3.3 STR and harmonic probability analysis ................. 134
4.4 Discussion .................................................................. 135

5 Assessment of the eTone Strategy .................................. 138
5.1 Introduction ............................................................. 138
5.2 General methods ...................................................... 141
  5.2.1 Subjects ............................................................. 141
Chapter 1: Introduction

1 Introduction

1.1 Overview

Significant advances in the development of cochlear implant (CI) technology over the last three decades have resulted in improved functional hearing for people with severe to profound deafness. Most postlingually deafened adults are now able to obtain some degree of speech comprehension (e.g., Balkany et al., 1996; Staller et al., 1997; Osberger and Fisher, 2001; Anderson et al., 2002). However, while high levels of speech perception can be achieved (e.g., Skinner et al., 1994; Garnham et al., 2002), the limited spectral and temporal information that is provided by existing CI systems appears to be insufficient for satisfactory music and tone perception in CI recipients and many report being unsatisfied with their ability to perceive musical sounds (e.g., Gfeller et al., 2000a; Leal et al., 2003). Subjects are generally able to perceive rhythmic aspects of music, however pitch and melody perception, as well as instrument identification and timbre perception, are more difficult (Gfeller and Lansing, 1991; Pijl and Schwarz, 1995; Leal et al., 2003). Furthermore, for speakers of tonal languages in which voice pitch (or tone) is used to convey lexical meaning, recognition of speech is adversely affected by poor pitch perception (Huang et al., 1996; Ciocca et al., 2002; Lee et al., 2002; Barry et al., 2002). Those studies have demonstrated poor tone discrimination by CI users for some tonal contrasts with results at, or marginally above, chance-level in many cases. Thus, improved coding and perception of pitch is of substantial importance to CI users, particularly speakers of tonal language. In the Asia Pacific region, there are an estimated 1.3 billion people who speak a tonally-based language, including those of China, Hong Kong, Taiwan, Vietnam, Thailand and Laos. CI users from those regions, particularly young children whom are developing language, along with their families and people that they communicate with, can all potential benefit from improved pitch coding systems.

While present CI systems are unable to encode fine spectral and temporal (spatio-temporal) information necessary to convey pitch as in normal hearing, timing information encoded through electrical pulse-rate can be used to convey a sensation of monophonic pitch, at least for pulse-rates up to approximately 300 Hz (e.g., Eddington et al., 1978; Pfingst 1988; Pfingst et al., 1994, Shannon 1983, 1993, Tong et al., 1983; Tong and Clark, 1985; Townshend et al., 1987; Busby and Clark 1997; Zeng, 2002). In
addition, for low-rate unmodulated pulse trains it has been shown that the relationship between pulse-rate and pitch in electrical hearing can be similar to that between F0 and pitch in acoustic hearing (e.g., Pijl and Schwarz 1995; McDermott and McKay 1997). That relationship forms an important basis for coding of F0 in CIs if normal musical intervals are to be encoded. Existing clinical sound coding strategies encode F0 temporal information by way of amplitude modulation in the stimulus envelope. Like pulse rate, changes in amplitude modulation rate of electrical pulse trains has also been shown to convey a sensation of pitch (e.g., Shannon, 1992; McKay et al., 1994, 1995; McDermott and McKay, 1997; Geurts and Wouters, 2001). However, those percepts can be weaker than those elicited by changes in pulse rate particularly for shallow modulation depths. Furthermore, coding of F0 modulation by current clinical strategies has been shown to be poor namely due to inconsistent presentation of deep modulation across channels and signals (e.g., Geurts and Wouters, 2001; McDermott, 2004; Vandali et al., 2005). Those issues may in part account for the unsatisfactory results of pitch and tonal language perception tests observed using clinical devices (e.g., Gfeller and Lansing, 1991; Huang et al., 1996; Pijl, 1997b; Ciocca et al., 2002; Lee et al., 2002; Leal et al., 2003; Looi et al., 2004, 2008), although the lack of fine-spatio temporal coding in CIs is also a major contributor.

A number of experimental strategies have been developed that enhance coding of temporal envelope cues to F0 (e.g., Geurts and Wouters 2001; Green et al., 2004, 2005; Vandali et al., 2005; Laneau et al., 2006; Mileczynski et al., 2009, 2012). Those strategies have been shown to provide significantly better pitch discrimination when compared to clinical strategies for a limited range of sounds and F0s (up to approximately 300 Hz). Those improvements are thought to be related to increased salience of rate-pitch information arising from the provision of deep and in-phase F0 modulation across stimulus channels. However, it is unclear from previous studies as to how accurately changes in F0, and hence changes in pitch-height, are represented in the auditory system to this form of stimulation (e.g., McKay et al., 1995; Busby and Clark 1997) and there is likely to be considerable variability across CI recipients. Thus further research is required to determine the effects of modulation shape on pitch height so that strategies which more accurately code F0 information can be developed. In addition, a likely consequence of applying deep F0 modulation in the stimulus envelope is that it will adversely affect coding of channel loudness (McKay and Henshall, 2010) and subsequently degrade perception of spectral and temporal envelope information needed
to convey speech information (e.g., Dawson et al., 2007). Thus determining the effects of deep modulation on channel loudness and how these effects can be compensated for in a strategy, if necessary, also requires further research. Furthermore, because pitch perception in CI recipients is thought to be derived from some weighted combination of rate and place of stimulation (e.g., McKay et al., 2000) changes in place of stimulation that are inconsistent with changes in F0 are likely to adversely affect judgement of pitch. Thus the influences of place-coding on rate-pitch must also be considered in the present research.

Improvement of pitch perception in CI recipients may also be gained through the use of training. Previous studies have shown that psychoacoustic training can improve frequency and pitch discrimination in non-musically trained normal-hearing (NH) listeners (e.g., Spiegel and Watson 1984, Micheyl et al., 2006). Micheyl and colleagues found that discrimination thresholds for non-musically trained listeners approached that of highly experienced musicians after approximately 4-8 hours of training. Effects of training on recognition of Mandarin tones by non-tonal-language normal-hearing listeners have also been observed (Wang et al., 1999). Wang and colleagues observed an average improvement in tone recognition of 21 percentage points after training which was retained even six months after training had ceased. Similarly, positive effects of training have been observed in CI recipients in music related tasks. For instance, improvements after training in instrument identification, melody recognition, and melodic contour identification have been observed (e.g., Gfeller, et al., 2002c; Galvin et al., 2007). In particular, Galvin and colleagues found that training using a melodic contour identification paradigm not only improved listeners’ ability on that task, but also generalised to improved identification of familiar melodies. However, while positive effects of training have been observed for stimuli in which consistent rate and place cues to pitch are provided, it has been demonstrated that when more natural musical sounds are used, changes in spectral timbre (resonant frequency or place of stimulation in CIs) that are inconsistent with F0 can strongly influence both normal-hearing (e.g. Pitt 1994) and cochlear implant listeners judgement of pitch (e.g., Sucher and McDermott, 2007; Galvin et al., 2008). That effect was shown to be particularly evident for listeners that were not musically trained which raises the possibility that musical training/experience might help listeners to discriminate pitch even when confounding effects of spectral timbre are present. Further research is however needed to examine whether training paradigms that specifically teach listeners to attend to F0 information (rate in CIs)
exclusively to judge pitch and resonant frequency (place in CIs) to discriminate spectral timbre may help them to judge those musical attributes independently when listening to real world complex musical sounds.

In summary, the outcomes of this research will provide useful information relevant to coding of temporal envelope cues to pitch using electrical stimulation. This information will facilitate the development of a real-time sound coding strategy that aids pitch perception without compromising coding and perception of segmental information necessary for perception of speech. In addition, this research will examine the effects of a training paradigm designed to help listeners perceive pitch and to better discriminate between cues to pitch and spectral timbre. Successful outcomes will be of significance to most CI users in terms of their ability to discrimination pitch, perceive melody, and appreciate music. In addition, listeners may benefit through improved perception of tone in tonal languages, intonation, word emphasis, gender and speaker identification, emotional state, and separation of concurrent speakers (which may perhaps aid speech perception in noise).
1.2 Statement of research questions

1.2.1 Aims

The primary goal of this research is to improve/optimise pitch perception for CI users. The research is divided into two parts. The first addresses methods of improving strategies that code rate-pitch information. The second explores methods of improving an individual’s ability to attend to this information. These aims are specifically defined as follows:

1) Improve pitch perception of monophonic harmonic sounds while maintaining speech recognition. This may be achieved through the development and evaluation of a strategy that enhances coding of F0 information in the temporal envelope of the stimulus signal but maintains coding of the spectral envelope and low-frequency temporal envelope information that is provided by existing strategies; and

2) Help listeners to better perceive pitch, particularly in music where many attributes of musical sounds vary together, through the development and evaluation of a training paradigm that teaches them to attend to F0 information exclusively for judgement of pitch and to resonant frequency information exclusively for discrimination of spectral timbre.

1.2.2 Hypotheses

For each aim a number of specific hypotheses were tested in the research:

1) Application of F0 modulation characterised by a sharp onset and rapid exponential decay, so that each F0 interval is primarily imparted by a single electrical pulse within each channel of stimulation will (a) produce more salient and accurate encoding of F0 rate-pitch information, than that provided by existing strategies; but will affect channel loudness when applied in a sound coding strategy and so (b) compensation for such effects may be needed so as to produce similar loudness to that produced by existing strategies and thereby maintain existing levels of speech perception.

2) Training involving discrimination of single cues to attributes of musical sounds (namely F0 for pitch and resonant frequency for spectral timbre) in the absence of other cue variations will: (a) improve listeners’ sensitivity to those cues; (b) improve sensitivity to F0 in complex sounds in which multiple cues vary; and (c) subsequent
training with complex sounds in which multiple cues vary may further improve listeners’ ability to discriminate the pitch of complex sounds.

1.3 Structure of thesis (chapter outline)

A brief description of the cochlear implant, clinical sound coding strategies, and signal properties related to percepts of pitch and timbre are provided in Chapter 2. This is followed by reviews of literature related to place- and rate-pitch perception in CI hearing, music and tonal language perception in CI hearing, pitch coding strategies for CIs, and training methods for improving auditory processing in normal-hearing and CI listeners.

Chapters 3 to 5, report on three studies conducted within the present research related to the first aim. Peer-reviewed journal articles were published for each of those studies during the course of the doctoral study (refer to Vandali et al., 2013; Vandali and van Hoesel, 2011; Vandali and van Hoesel, 2012, respectively). Because the content of those chapters was derived almost verbatim from the journal articles, there is some overlap between chapter introductory material and the literature review presented in chapter 2.

In chapter 3, results of a series of psychophysical experiments are reported that examined parameters of amplitude-modulated electrical pulse trains to determine their effect on salience and accuracy F0 rate information in electric hearing. In those experiments, the effects of amplitude modulation depth, rate, and shape, and overall presentation level, on rate-pitch and loudness were examined. Based on the results of the psychophysical experiments and on previous research, an experimental rate-pitch coding strategy was developed to enhance coding of F0 modulation information in the stimulus envelope. The strategy, known as Enhanced-Envelope-Encoded Tone (eTone), was designed to provide a more salient and accurate representation of F0 rate information than that provided by existing strategies. A complete description of eTone is provided in chapter 4 along with results of laboratory tests examining performance of critical processing stages, namely the F0 estimator and harmonic probability estimator.

In chapter 5, results of a perceptual study with CI recipients comparing eTone to a clinical coding strategy (ACE) are presented. In that study, pitch discrimination is compared between strategies for synthetic complex harmonic tones and natural sung vowels. Speech recognition in quiet and noise is also compared.
Chapter 1: Introduction

Chapters, 6 and 7 report on the second aim of the research. In chapter 6, a musical-pitch training program developed within the present doctoral study is described. The program was designed to initially teach listeners to attend to single-isolated acoustic cues, namely F0 for pitch or resonant frequency for spectral timbre, and later to the use of multiple concurrent cues for perception of musical attributes in more complex sounds. The effects of that training program on pitch and spectral timbre discrimination by normal-hearing listeners and cochlear implant recipients are presented in chapter 7. For the CI users, pitch perception in the presence of varying spectral timbre was also examined. Follow up tests were also conducted with the CI recipients some time after training had ceased to examine the robustness of any training effects. At the time of writing this thesis, some of the content of chapters 6 and 7 had been submitted as a journal article (Vandali et al., submitted).

The final chapter discusses the outcomes of each study and their combined relevance to the aims and hypotheses of the research.
Chapter 2: Background and Relevance to Research

2 Background and Relevance to Research

2.1 Cochlear implants and sound coding strategies

Most modern cochlear implant systems comprise three components, an external wearable sound processor, an implanted receiver/stimulator, and an array of intra-cochlea electrodes, see Fig. 2.1. See Zeng (2004) and Wilson (2004) for reviews. The sound processor generally consists of: a microphone to measure the sound signal; an electronic processor, generally in the form of a digital signal processor (DSP), which is used to encode the sound signal into appropriate stimulus signals; and a radio-frequency transmitter which is used to send the encoded stimulus signals transcutaneously to the implanted receiver/stimulator via a transmitting coil. The receiver/stimulator consists of a surgically implanted electronic device which receives and decodes the stimulus signals from the sound processor and generates corresponding electrical signals for activation of the intra-cochlea electrodes. Small electrical current pulses are passed between the electrodes to stimulate neurons within the peripheral auditory system thereby eliciting hearing sensations.

Figure 2.1. Components of a cochlear implant system. A sound processor and microphone are located behind the ear, although in earlier systems the sound processor was larger and was typically located elsewhere on the body such as in a pocket or on a belt. The stimulus signal generated and encoded by the sound processor is transmitted transcutaneously via a transmitting coil placed above the receiver/stimulator which is surgically implanted on the mastoid bone. The receiver/stimulator receives and decodes the stimulus signal and generates corresponding electrical current pulses that are delivered to the electrodes implanted within the cochlea.
Chapter 2: Background and Relevance to Research

The sound processing strategies that are presently available in clinical systems perform a spectral (frequency) analysis of the signal and from that generate corresponding electrical stimuli to deliver to the receiver/stimulator. See Loizou (1998) for a review. Essentially they divide the incoming sound signal into a number of frequency channels using a bank of band-pass filters, see Fig. 2.2. The temporal envelope of the signal in each channel is usually estimated, although some strategies utilise the band-pass signals directly. Electrical stimuli are derived from these channels and are used to activate corresponding electrodes within the cochlea using charge-balanced, biphasic current pulses presented at a constant stimulation rate, variable rate, or at different fixed rates across electrodes depending on the strategy. In earlier strategies, analogue rather than pulsatile stimulation was used. Electrodes are allocated to frequency channels in a manner consistent with the tonotopic arrangement of the cochlea. The two most commonly used clinical strategies presently available are the Continuous Interleaved Sampling (CIS) strategy (Wilson et al., 1991) and the Advanced Combinational Encoder (ACE) strategy (Vandali et al., 2000; Skinner et al., 2002). In general those strategies code spectral information in the signal via place of electrical stimulation whereas temporal and/or temporal envelope information is coded via rate of stimulation or via amplitude fluctuations in the envelope of the electrical signal respectively. The spectral and temporal information encoded by existing cochlear implant systems has been shown to provide sufficient information to allow CI recipients to understand speech without the aid of lip reading (e.g., Van Tassel et al., 1987, 1992; Skinner et al., 1994; Balkany et al., 1996; Staller et al., 1997; Osberger and Fisher, 2001; Anderson et al., 2002; Garnham et al., 2002; Xu and Pfingst, 2008). However, perception of voice and musical pitch has been shown to be far from satisfactory compared to that of normal-hearing listeners (Leal et al., 2003; Gfeller and Lansing, 1991; Pijl and Schwarz, 1995; Huang et al., 1996; Gfeller et al., 2000a; Ciocca et al., 2002; Lee et al., 2002; Barry et al., 2002; Leal et al., 2003) and improvement of its perception is thus one of the main aim of the present research.
Chapter 2: Background and Relevance to Research

Figure 2.2. Block diagram of a clinical sound coding system. From left to right, the incoming acoustic signal is measured using a microphone and amplified to a suitable level for further processing. For a DSP-based processor, the signal is then digitised via an analogue-to-digital converter. The signal is then spectrally analysed into a number of frequency channels using a bank of band-pass filters and envelope detectors. In the ACE strategy, a subset of the channels signals are selected for further processing based those having the largest amplitude at any one time. In the case of the CIS strategy, all channels are used in further processing. The level of the channel signals are converted to an electrical stimulation level that falls within the electrical dynamic range of measured threshold and comfortable-loudness stimulation levels for each electrode that is allocated to each of the channels. The stimulation levels for each electrode, together with other parameters of the stimulus, such as phase-duration, mode of stimulation and rate of stimulation, are encoded into a format that can be recognised by the implanted receiver/stimulator and are transmitted transcutaneously to the implant as a radio-frequency (RF) signal via the transmitting coil.

2.2 Properties of musical and voiced tones

Musical instruments and the human voice are capable of producing tonal (voiced) and non-tonal (unvoiced) sounds. Tonal sounds exhibit periodic structure (i.e., repeated patterns of vibration) in their acoustic waveform whereas non-tonal sounds have an aperiodic structure (random non-repeating pattern of vibration). The simplest tonal sound is a pure tone consisting of a single frequency component. Most tonal sounds encountered in everyday life are composed of multiple frequency components and are known as complex harmonic tones. The lowest frequency component in a complex harmonic tone is known as the fundamental frequency (F0). Additional frequency components forming a complex harmonic tone are whole-number multiples of F0 known as harmonics. For both pure tones and complex harmonics tones, a sensation of pitch may be quantified according to the fundamental frequency. Other acoustic properties of sounds give rise to perceptual attributes such as loudness (based on the intensity of the sound) and timbre (based mainly on properties other than intensity and fundamental frequency such as spectral resonance(s) and temporal envelope).
Chapter 2: Background and Relevance to Research

2.2.1 Pitch

“Pitch” as used throughout this manuscript refers to “the attribute of auditory sensation that can be ordered on a musical scale from low to high” (ASA 1960, ANSI 1994). It is an attribute of sound that can be used to produce melodies. For harmonic sounds, it represents the perceived fundamental frequency of the sound. It is one of the major auditory attributes of musical tones along with duration, loudness, timbre, and sound source location. It is multidimensional, at least involving the components of pitch height and pitch strength (or salience). Pitch strength can be defined as the certainty of a pitch height judgment, or alternatively, the variance of pitch height judgments represent a measure of pitch strength (Beerends and Houtsma 1989; Rakowski, 1996).

Sequences of sounds in which pitch varies create melody and thus pitch perception is integral to one's appreciation of music. In speech, pitch pertains to the perceived fundamental voicing frequency (F0) of the talker. It plays an important role in speech perception as it provides cues to linguistic features such as intonation and word emphasis (Highnam and Morris, 1987; Wells et al., 1995), to perception of paralinguistic features such as gender, speaker identification, and the emotional state of the speaker (Abberton and Fourcin, 1978; Liberman and Michaels, 1962), and to segregation of speech from concurrent speakers (e.g., Assmann and Summerfield, 1990; Brokx and Nootebroon, 1982). Voice pitch information is also important to perception of tonal languages, such as Mandarin and Cantonese, where a change in F0 within the same phonemic segment causes a change in lexical meaning (Gandour, 1981).

2.2.2 Timbre

Timbre, also known as tone colour, is the quality of a musical sound or tone that distinguishes types of sound production, such as different voices and musical instruments. It is “an attribute of an auditory sensation which a listener uses to judge that two sounds similarly presented and having the same loudness and pitch are dissimilar. Timbre depends primarily upon the spectrum of the stimulus, but it also depends upon the waveform, the sound pressure, the frequency location of the spectrum, and the temporal characteristics of the stimulus” (ASA 1960, 45). In the context of the present research, “spectral timbre” refers to the aspect of timbre derived from the spectrum of the stimulus in which a percept can be ordered on a scale of “dull” to “bright”.

11
2.3 Pitch coding in cochlear implants

Extensive research over the last three decades examining the perception of pitch using electrical stimulation has shown that place of stimulation can convey a sensation of pitch according to a tonotopic arrangement where stimulation of neurons adjacent to basal electrodes (see Fig. 2.3) generally elicit a higher pitch than those for apical electrodes (e.g., Tong and Clark, 1985; Townshend et al., 1987; Busby et al., 1994; McDermott and McKay 1994; Nelson et al., 1995). This percept is typically referred to as place-pitch.

![Figure 2.3. Cross section of cochlea with inserted electrode array. Stimulation of auditory neurons adjacent to electrodes near the base of the cochlea elicit a higher (place) pitch percept whereas those closer to the apex elicit a lower pitch percept.](image)

Rate of stimulation, or modulation rate coded using a higher constant stimulation rate (see Fig. 2.4), has also been shown to elicit a percept of pitch referred to as rate-pitch (e.g., Shannon, 1983; Tong and Clark, 1985; Townshend et al., 1987; Shannon, 1992; McKay et al., 1994, 1995; Zeng, 2002). It is thought that some weighted combination of rate and place of stimulation is utilised by CI recipients (to different degrees) to judge pitch (e.g., McKay et al. 2000). However, each mechanism exhibits a limited capacity to encode information necessary for normal pitch perception. For instance, for place of stimulation, the excitation field resulting from the stimulus current cannot be sufficiently focused along the basilar membrane to recruit a local population of auditory neurons...
Chapter 2: Background and Relevance to Research

(Townshend et al., 1987; Cohen et al., 2003, Cohen, 2009) and hence the spectral, or place-pitch, selectivity encoded by existing devices is significantly reduced as compared to that available in normal hearing. In addition, the normal relationship between place of stimulation and frequency is not necessarily preserved due to sub-optimal position of electrodes relative to surviving auditory neurons, and variable/poor neural survival (Busby et al., 1994; Cohen et al., 1996). For rate of stimulation, the ability to perceive changes in rate deteriorates markedly for rates above approximately 300 pulses-per-second (PPS) in general (e.g., Shannon, 1983; Tong and Clark, 1985; Townshend et al., 1987; Zeng, 2002). In addition, sensitivity to differences in electrical stimulation rate is generally significantly worse than similar measures observed in normal hearing.

![Unmodulated pulse-train](image1)

![Modulated pulse-train](image2)

**Figure 2.4.** Unmodulated and amplitude-modulated biphasic current pulse trains. Bi-phasic current pulses consist of two charge-balanced phases of equal phase-width and equal but opposite current level that are separated by an inter-phase gap. Rate-pitch can be elicited by changes in the pulse-rate of an unmodulated pulse-train or by changes in the modulation rate of an amplitude-modulated pulse-train.

In normal hearing, the pitch of a complex tone, such as a voiced vowel or musical tone, is primarily derived from fine spectral and temporal detail and/or spatio-temporal information that is encoded in the peripheral auditory system by auditory filters in which low-order harmonics of the fundamental frequency are resolved (Ritsma, 1967; Moore et al., 1985; Dai, 2000), see Fig. 2.5.
Chapter 2: Background and Relevance to Research

Figure 2.5. A schematic representation of the transduction process for a voiced vowel from the acoustic input, through to basilar membrane vibration in the cochlea. The instantaneous spectrum of the vowel is shown in the top panel. A schematic representation of the auditory filterbank resulting from the mechanical properties of the middle ear and cochlea is shown next followed by the spectral excitation pattern and the basilar membrane vibration pattern over time to the incoming vowel. The fundamental frequency (F0) and low-order harmonics of F0 are resolved by the auditory filters as can be seen by the peaks occurring at multiples of F0 in the excitation pattern and by the sinusoidal vibration on the basilar membrane that corresponds to the frequency of each resolved harmonic. Higher order harmonics of F0 are not resolved by auditory filters and only the broad spatial (spectral) envelope of the signal is captured in the excitation pattern. Basilar membrane vibration to those unresolved harmonics consists of the combination of multiple harmonics and F0 information is represented in the temporal envelope of vibration.

Given the inability to encode fine spectral detail and high temporal rates using electrical stimulation, it comes as little surprise that normal pitch cannot be conveyed by present
CI systems. However, normal-hearing listeners can also derive pitch from temporal envelope information encoded by auditory filters in which F0 harmonics are not resolved (Houtsma and Smurzynski, 1990) or in which fine spectral and temporal details are not available (e.g., Burns and Viemiester, 1976, 1981). That mechanism is thought to be closely related to that of rate-pitch perception in electrical stimulation (McKay and Carlyon, 1999) and exhibits similarly poorer F0 discrimination when compared to pitch derived from fine spatio-temporal structure in normal hearing (Shackleton and Carlyon, 1994). Developers of CI sound coding strategies have thus focused on employing rate, or modulation rate, to convey information about the temporal envelope (e.g., prosody and F0 information up to approximately 300 Hz). In addition, as will be explained in forthcoming sections, the limited spectral resolution and selectivity provided by place of stimulation in present CI devices, means that for low F0s within the typical voicing frequency range, place coding can not be used to convey accurate/useful information about F0 harmonics needed for perception of voice and musical pitch. Instead present CI systems utilise place coding to convey information about the spectral envelope (e.g., vowel formant frequencies and spectral timbre of musical instruments). For high F0s, beyond the typical voicing frequency range, place coding may be sufficient to convey some spectral cues to F0 and hence pitch (e.g., Vandali et al., 2005; Laneau et al., 2006). However, because changes in place of stimulation can affect listeners’ judgement of pitch, problems arise when those changes are not consistent with changes in F0. For example, an increase in formant frequencies across two vowels may be heard as a rise in pitch even when F0 decreases across the vowels. Thus, while the present research is concerned mainly with improving coding of temporal/rate cues to F0 pitch, a review of place-pitch is also warranted so as to provide a better understanding of the problem(s).

While it is possible that future implantable devices may have the capacity to deliver more detailed spectral and temporal information, when and if this technology becomes available is unclear. Therefore, as there are a substantial number of people that can presently benefit from improved coding of pitch, further research into methods of enhancing envelope rate-pitch perception using present-day implantable devices is warranted.
Chapter 2: Background and Relevance to Research

2.3.1 Place-pitch

2.3.1.1 Place-pitch resolution

Studies examining pitch discrimination between stimuli presented at different electrode places, while holding rate of stimulation fixed, have shown considerable variability in place-pitch order and sensitivity across subjects. Nelson et al., (1995) examined the ability of fourteen Nucleus 22 CI users to distinguish electrical stimuli presented at different electrode places on the basis of “pitch” or “sharpness”. Performance in an electrode pitch ranking task, defined in terms of $d'$-prime ($d'$) per distance along the cochlea, was measured (where $d'$ is a measure of dissimilarity with values greater than 1 indicating a statistical significant difference, e.g., MacMillan and Creelman, 1991). Subjects with good to excellent place-pitch resolution obtained ceiling level performance and typically ranked all electrodes in correct tonotopic order according to their place. For moderate to poor performers, some instances of reversals in place-pitch ordering were observed on some electrodes. D-prime sensitivity between electrode places varied across subjects from 0.12 to 3.16 $d'/mm$ (or 0.09 to 2.37 $d'/electrode$ for an electrode spacing of 0.75 mm which is typical of Nucleus CI devices). In a review of this data, Moore and Carlyon (2005) observed that the median discrimination threshold was about 1.2 mm (or 1.6 electrode places) which according to Greenwood’s equation of frequency versus place (Greenwood, 1990), corresponds to an acoustic frequency change (i.e., frequency difference limen – FDL) of approximately 21%. The best performer had a threshold of around 0.25 mm corresponding to a FDL of 3%. Other researchers have reported similar low thresholds ranging from 0.25 to 0.46 mm in a small group of Nucleus 24 CI recipients (Laneau and Wouters, 2004a). While those thresholds are low compared to median thresholds reported Moore and Carlyon (2005), they are substantially worse than those of normal-hearing listeners who can obtain FDLs as low as 0.2% for a 1 kHz pure tone (Moore, 1973).

2.3.1.2 Place-pitch selectivity

The pattern of spatial excitation in the cochlea to electrical stimuli provides a basis by which place coding selectivity can be gauged. Spatial excitation patterns have been measured using a number of different techniques. For instance, psychophysics forward masking tests in which probe threshold as function of spatial distance from a fixed-place masker have been employed to estimate the pattern of excitation to the masker (e.g.,
Cohen et al., 1996, 2003; Chatterjee and Shannon, 1998). Results have shown that masking (and hence the excitation pattern) is generally greatest at sites nearest the masking electrode and decays at more distant sites. Cohen et al., (1996) demonstrated that the forward masking patterns were broadly consistent with place-pitch estimation patterns. Non-subjective tests, such as neural response telemetry, have also been used to measure the “effective stimulation field” produced by electrical stimuli (e.g., Cohen et al., 2003; Cohen 2009). Outcomes from those studies have shown that excitation patterns to bipolar and monopolar stimulation are typically quite broad in shape with a peak at the site closest to the stimulating electrode. Cohen (2009) modelled the effective spread of current within the fluid and tissue surrounding the stimulating electrode using a function in which current decayed logarithmically (or linearly in clinical units for the Nucleus CI system) from the site of stimulation along the spatial dimension of the cochlea. Spread functions for five Nucleus 24 subjects displayed an average current decay rate of approximately 1 dB (or 6 clinical units where each unit is a step 0.176 dB current) per electrode place (or 0.75 mm which is the distance between electrodes in the Nucleus CI system). Given that the typical electrical dynamic range for CI users is around 30-45 clinical units (i.e., 5-8 dB current), these data imply that neurons adjacent to electrodes many sites away from the stimulating electrode (e.g., 5-8) could be activated.

Place-pitch selectivity has also been examined by measurement of subject’s ability to resolve spectral peaks (e.g., Henry et al., 2003, 2005; Won et al., 2007). Henry et al., (2003) used stimuli consisting of spectral ripple shaped noise and measured discrimination thresholds for a reversal in the phase of the rippled shape. Thresholds were measured as a function of ripple density and number of channels used in CI recipient’s processors. Thresholds for normal-hearing listeners using a channel vocoder simulation of CI processing were also measured. They found that spectral resolution thresholds for CI users plateau at around 4-6 channels in contrast to approximately 16 channels for simulated CI processing in NH listeners (which was somewhat consistent with outcomes of studies examining speech recognition as a function of channels by CI recipients and for CI simulated speech by NH listeners, e.g., Shannon et al., 1995; Dorman et al., 1998). For 12 channels, average spectral resolution thresholds were around 3000 Hz for CI recipients compared to 400 Hz for simulated CI processing in NH listeners. Henry et al., (2005) went on to demonstrate a relationship between spectral peak resolution and vowel and consonant recognition in quiet. They also found
that the degree of spectral peak resolution for accurate vowel and consonant recognition was around 4 ripples/octave and that a resolution less than 1-2 ripples/octave (which is typical of CI users in which an average threshold of 0.62 ripples/octave was measured) could result in poor speech recognition. Won et al., (2007), also demonstrated that spectral ripple thresholds were correlated with speech reception thresholds for word recognition in quiet and in noise.

### 2.3.1.3 Place-pitch height

Blamey et al. (1996) compared pitch-height of acoustically presented pure tones to electrical pulse trains presented at fixed electrode sites and fixed rates of stimulation in CI users who had some residual hearing in their non-implanted ear. They found that both electrode place and pulse rate influenced perceived pitch and possibly elicited two perceptually distinct components. The pitch arising from variation in electrode place (with pulse rate fixed) was much lower than that predicted by Greenwood’s (1990) formula relating characteristic frequency to cochlea place. Similar findings have been observed in more recent studies (e.g., Boëx et al., 2006; Dorman et al., 2007). Blamey and colleagues attributed the findings to several possibilities which included: activation of spiral ganglion cells in Rosenthal’s canal that correspond to cochlear sites more apical than those of the electrode sites; adaptation to sounds delivered simultaneously by the two modes of stimulation so that the perceived pitch was the same in both ears; a shift in the place of maximum excitation that occurs in a severely impaired ear; and the influence of the low-pitched component corresponding to pulse rate of the electrical stimuli on the overall pitch sensation. However, another possibility was proposed by Carlyon et al., (2010b) who suggested that those earlier findings may be due to non-sensory biases and procedural effects and that place-pitch can be well accounted for by the predictions of Greenwood’s formula. In addition, McDermott et al., (2009) observed pitch matches closer to those predicted by Greenwood’s model in a group of five newly implanted bimodal hearing subjects.

### 2.3.1.4 Inter-electrode place-pitch

Some researchers have observed that activation of two adjacent electrodes, either simultaneously or in quick succession, can produce an intermediate place-pitch percept to those produced by each electrode separately (Townshend et al., 1987; McDermott and McKay, 1994; Donaldson and Krefl, 2005). In general, the intermediate place-pitch is
determined by the ratio of stimulation current delivered to the two electrodes. McDermott and McKay (1994) activated two electrodes in quick succession and found that an intermediate place-pitch could be obtained between adjacent electrode pairs in half of the subjects tested. They suggested that the effect was caused by the spatial overlap of the excitation patterns produced by activation of each electrode. Donaldson et al., (2005) activated electrodes simultaneously and reported discrimination thresholds for single versus dual-electrode stimulation ranging from 0.11 to 0.64 electrode places (or mm, for the electrode spacing of 1 mm in the Clarion CII device ) reflecting a two to ninefold increase in the number of place-pitch steps possible with dual-electrode stimuli. Expressed in terms of acoustic frequency differences, those data correspond to FDLs of approximately 1.3 to 10% which are still somewhat worse than those observed in normal hearing (§2.3.1.1).

2.3.2 Rate-pitch

Increases in rate of electrical stimulation can be used to convey a sensation of increasing pitch for rates as low as 50 PPS and up to 300 PPS on average (Eddington et al., 1978; Pfingst 1988; Pfingst et al., 1994, Shannon 1983, 1993, Tong et al., 1983; Tong and Clark, 1985; Townshend et al., 1987; Busby and Clark 1997; Zeng, 2002). In addition, some studies have reported that further rate increases up to as high as 1000 PPS can be detected by some subjects (e.g., Townshend et al., 1987; Wilson, 1997a; Landsberger and McKay, 2005), although those results are not necessarily based on discrimination of changes in pitch and a rate of around 300 PPS is generally accepted as the upper limit of rate-pitch perception. As a consequence, the full range of fundamental frequencies produced by different speakers, which can exceed 300 Hz for some women and children and which can extend to well over 1000 Hz for some musical instruments, cannot be conveyed using electrical rate and so the application of this research must exclude those high voicing frequencies and musical notes just above middle C in the western musical scale.

Experiments conducted in normal-hearing listeners in which place and rate information are presented independently have shown a higher upper limit, of around 600 PPS, in rate-pitch discrimination compared to that seen in CI users. Carlyon and Deeks (2002) measured F0 discrimination using band-pass filtered complex harmonic signals that resembled filtered pulse trains. When the harmonics were filtered between 3900-5400 Hz, and summed in sine phase so that place and rate information were congruent,
Chapter 2: Background and Relevance to Research

subjects were able to correctly discriminate all F0 intervals tested up to 600 PPS. However, when harmonics were summed in alternating phase so that rate information was doubled compared to information provided by harmonic place, discrimination thresholds increased with increasing F0 until the task became impossible at a pulse rate of 600 PPS. Those outcomes suggest that the inability to follow rates beyond 300 PPS in electric hearing are not entirely due to a central pitch limitation, because normal-hearing listeners were able to follow substantially higher rates when provided with analogous acoustic stimuli.

2.3.2.1 Neural mechanism underlying rate-pitch
Pitch percepts arising from temporal stimulus properties presumably reflect differences in neural response properties. Those differences are manifest in terms of the pattern of spatial activation (§2.3.1.2), which for electrical stimulation cannot be sufficiently focused along the basilar membrane to recruit a local population of auditory neurons (Townshend et al., 1987; Cohen et al., 2003, Cohen, 2009), and in terms of their temporal response. For electrical stimulation at low rates of less than a few hundred pulses-per-second (and up to approximately 800 PPS in some cases), a very deterministic response behaviour with a high degree of phase locking and entrainment is observed in single unit responses as has been shown in auditory nerve data from animal (Kiang et al., 1979; Javel et al., 1987; Van den Honert and Stypulkowski, 1987; Hartmann and Klinke, 1990; Sly et al., 2007) and human studies (Wilson et al., 1997b; Matsuoka et al., 1998). This contrasts with acoustic stimulation in which a more stochastic response behaviour to acoustic pulses is observed in single units but a higher degree of phase locking in the population (ensemble) response is seen for rates up to 4-5 kHz (e.g., Rose et al., 1968; Kiang et al., 1965; Moore, 1973; Micheyl et al., 1998; Moore and Sek, 1996). At higher rates of electrical stimulation, where relative refractory effects become more dominant (800 to 2000 PPS), entrainment decreases and population responses exhibit poorer phase locking than that seen at lower rates to each pulse in the stimulus (Parkins, 1989; Dynes and Delgutte, 1992). That behaviour arises because immediately after firing, neurons are prevented from discharging again during an absolute refractory period, which is followed by a period of relative refraction in which their firing probability is gradually restored (e.g., Shepherd and Javel, 1997; Miller et al., 2001; Shepherd et al., 2004).
Chapter 2: Background and Relevance to Research

2.3.2.2 Rate-pitch difference limens

In a review of five studies (Pfingst et al., 1994; van Hoesel and Clark, 1997; McKay and Carlyon 1999; McKay et al., 2000; Zeng 2002) comprising a total of 19 CI subjects, Moore and Carlyon (2005) reported an average electrical rate-pitch difference limen (DL) of 7.3% (with a range of approximately 2 to 18%) for a nominal pulse rate of 100 PPS. The large variability in performance across subjects, which is typical of many psychophysical measures in CIs, may stem from differences in procedures across studies (such as whether loudness roving was employed when determining DLs), and differences in subjects (such as age, duration of deafness, previous pitch/music experience, pathology, auditory neural survival and proximity of stimulating electrodes to neurons). Nonetheless, those data compare poorly to normal-hearing listeners where frequency DLs of less than 1% (down to 0.15%) are typically for pure tones (Moore, 1973) and F0 DLs of around 0.25% are typical for complex harmonic tones (Houtsma and Smurzynski, 1990). Those data suggest that the salience (or strength) of rate-pitch information coded electrically is somewhat weaker than cues to pitch provided by acoustic stimuli in normal hearing in which spatio-temporal information is available. The implication of those data to perception of music, where the smallest interval (in western music) is one semitone (or 5.9%), is that we might expect only good to excellent CI performers to be able to accurately follow melody using rate information, assuming that F0 rate information is coded in a salient and reliable manner. Similarly, in tonal languages such as Mandarin and Cantonese, F0 intervals between some tones can be as low as 1-2 semitones (Bauer and Benedict, 1997) and thus again only good performers may be able to correctly discriminate all lexical tones from one another based on F0 rate information.

2.3.2.3 Rate-pitch height and scaling

Electrical psychophysics studies examining pitch intervals perceived by changes in pulse rate have shown that the relationship between pulse-rate and pitch in electric hearing is consistent with that between F0 and musical intervals in normal hearing, at least for low rates of stimulation (e.g., Pijl and Schwarz, 1995; McDermott and McKay, 1997). Pijl and Schwarz (1995) asked CI subjects to indicate whether the pitch interval between two sequentially presented stimuli that differed only in pulse-rate (i.e., stimulus electrode and level remained fixed) were “in tune” (consistent), “too low” (flat), or “too high” (sharp) with their memory of a specific musical interval. Musical intervals from tone sequences
Chapter 2: Background and Relevance to Research

in well-known melodies were used which included a minor 3rd, 4th, and 5th, and a major 6th. Pulse rate ratios between the two presented stimuli ranged from 1 to 12 semitones with the lower rate in the pair remaining fixed at 137, 145, 163, and 127 PPS for the four intervals respectively. For the three subjects tested, their judgement of musical intervals to changes in pulse rate appeared to be consistent with the same ratios that normal-hearing listeners would associate with changes in F0. Similar results were observed when the experiment was repeated at different electrode sites, although the subjects reported that apical stimulation produced more musical sounds. Similar results were also observed by McDermott and McKay (1997) who examined both musical interval perception and production in a single CI subject that had trained as a musical instrument tuner prior to deafness. In addition, for that subject, when sinusoidal amplitude-modulated pulse trains were presented which varied in modulation rate rather than pulse rate, similar results were also reported. Pijl and Schwarz (1995) repeated their experiment but randomly transposed the pulse rate of the lower rate tone to a lower octave or to higher octaves. Results for stimuli transposed to a lower octave were consistent with their original experiment however; the results for stimuli transposed to higher octaves exhibited a downward trend in perceived interval size of 1.2 semitones per octave increase in pulse-rate. That finding suggested that the correspondence between perceived pitch intervals and pulse-rate in CIs may apply only to a limited range of low rates.

2.3.2.4 Comparison of acoustic pure tone and electric pitch-height

While previous psychophysics studies using low pulse rates have shown that the relationship between pulse-rate and pitch in CIs is similar to that between F0 and pitch in NH, it is not clear from whether absolute pitch is similar for electric and acoustic rates. As discussed earlier (§2.3.1.3), Blamey et al. (1996) compared the pitch of acoustically presented pure tones to electrical pulse trains presented at fixed electrode sites and fixed rates of stimulation in CI users who had some residual hearing in their non-implanted ear. Results demonstrated that both pulse rate and electrode place influenced perceived pitch and possibly elicited two perceptual distinct components consistent with other studies (e.g., Tong et al., 1983; McKay et al., 2000). The average pulse rate of the matched electrical stimuli (when electrode place was fixed) were approximately equal to the frequency of the pure tones, although pitch increased more rapidly as a function of pulse rate than as a function of frequency over the range of rates
examined in the experiment (i.e., from 62 to 1000 Hz) and was higher for stimuli presented at a basal electrode compared to an apical electrode site. However, the range of pulse rates over which pitch varied was mainly restricted to rates below approximately 300 PPS and was consistent with a reduction in the salience of rate-pitch with increasing pulse rate. At higher pulse rates where the salience of rate-pitch was weak, the results could be explained by the effect of electrode place and can account for the higher pitch reported by subjects to the electrical stimuli than that predicted according to pulse-rate alone.

2.3.2.5 Effects of stimulation place on rate-pitch height

Direct effects of electrode place on rate-pitch have been observed (e.g., Pijl, 1997a; Fearn, 2001; Zeng, 2002). Pijl (1997a) conducted pitch matching experiments using low-rate (up to around 500 PPS) electrical pulse trains which differed in place of stimulation in two Nucleus CI22 implantees using bipolar stimulation. In general for low pulse rates (≤ 100 PPS), results demonstrated little effect on pitch to differences in place of stimulation (of +/-6 and 12 electrodes) although some variability across subjects was observed. For higher pulse-rates, the effects of place of stimulation were more significant and in general basalward shifts in place of stimulation produced a higher pitch percept. Zeng (2002) employed a pitch magnitude estimation test to examine the effects of place and rate of unmodulated pulse trains on pitch in four CI recipients. In general for low pulse rates, pitch was dominated by rate and little effect of place was observed. Pitch increased monotonically with increasing pulse rate up to approximately 300 PPS and saturated at higher rates. At those higher rates, pitch was dominated by place and more basal stimulation elicited higher pitch. The results of those studies are consistent with the notion that a sensation of pitch is derived from the place (or apical edge) of the spatial excitation pattern, particularly for rates approaching and higher than the upper limit of rate discrimination where the salience of rate-pitch cues are presumably weak.

2.3.2.6 Effects of stimulation level on rate-pitch height

It has been demonstrated that, like pitch, loudness generally increases with increasing pulse rate (e.g., Eddington et al., 1978, Pfingst et al., 1994, Shannon 1983, Tong et al., 1983). This raises the possibility that the pitch elicited by pulse rate may not be independent of stimulation level. This possibility has been verified in several studies.
although considerable differences in outcomes between subjects and studies have been reported (Pijl, 1997a; Arnoldner et al., 2008; Carlyon et al., 2010a). In the same study by Pijl (1997a) discussed above, the pitch of electrical pulse trains which differed in stimulation level was examined in a group of three Nucleus CI22 implantees. Results showed that unlike loudness, pitch decreased with increasing stimulation level and that this was particularly the case when the rate-pitch percept was weak (i.e., at higher pulse rates). It was postulated that the results were due to an apicalward shift (i.e., to a lower frequency) in the apical edge of the spatial excitation pattern with increasing stimulation level. That result is somewhat in agreement with data from normal-hearing listeners in which an apicalward shift in the position of maximum excitation occurs with increasing pure-tone amplitude (Zwislocki, 1991). Zwislocki pointed out however that this shift does not produce a corresponding change in pitch and suggested that the apical edge rather than the peak of the excitation region may be used to determine pitch.

Arnoldner et al. (2008) examined the effects of stimulation level on pitch in a group of sixteen Nucleus CI24 users and observed a significant effect of level on pitch in approximately 87% of subjects with 73% of them indicating a lower pitch with increasing stimulation level and the other 27% reporting the opposite. More recently, Carlyon et al. (2010a) observed that small changes in level can have substantial effects on pitch which are not entirely consistent with the effects of stimulation level on spatial excitation patterns. Using a pitch ranking tests, results in 16 out of 21 cases in which an effect of level was observed were consistent with pitch increasing with increasing level and the authors suggested that a temporal mechanism, in which neural discharge rates increase with increasing level, might operate.

2.3.2.7 Impact of place and level effects on rate-pitch

Given that both place and level of stimulation can affect rate-pitch, then coding of pitch using F0 rate information in a sound coding strategy may be adversely affected by changes in the spectrum and intensity of signals that take place independent of F0. This is a common occurrence in the real world where spectral information and intensity vary greatly across different vowels, speakers, and musical instruments. It can also arise for tones produced by the same instrument or by the same speaker of a vowel. For instance, trained singers who are conditioned to produce a balanced vocal quality (or “resonance”) can introduce variations in spectral timbre (e.g., in the amplitude distribution of harmonics and hence the mean location of formant frequencies) when singing at
different F0s (e.g., Joliveau et al., 2004). Such effects have been noted through spectral and electrodogram (i.e., the distribution of electrical stimulus pulses across electrodes and time) analysis of sung vowel stimuli (Looi, 2008; Vandali, 2008). Similar confirmation has been obtained by examining the spatial centroid (Laneau et al., 2004b, 2006) of stimulation derived from the mean location of the distributed stimulation pattern (Vandali, 2008; Swanson, 2008). Vandali and colleagues found that for the same singer and vowel produced at different F0s, the spatial centroid varied independently and often contrary to changes in F0. Poor outcomes for some subjects in pitch discrimination involving those stimuli were observed. Those data supported the notion that subjects were using place cues in addition to rate to judge pitch, consistent with outcomes of other studies (e.g., Tong et al., 1983; McKay et al., 2000; Fearn, 2001).

2.3.2.8 Effects of spatial interactions on rate-pitch

Studies examining the spread of neural excitation along the cochlea to electrical stimuli (e.g., Chatterjee and Shannon, 1998; Cohen et al., 2003, Cohen, 2009) have shown that the spatial excitation patterns from adjacent electrodes can overlap significantly and hence limit encoding of independent temporal information at nearby places. This limitation has also been born out in psychophysics studies examining integration of temporal information as a function of electrode separation. McKay and McDermott, (1996) examined the integration of low-rate temporal information coded on neighbouring electrodes as a function of electrode separation. In general they observed that modulated signals presented at electrodes separated by no more than 3 to 4 mm were perceived as the combined temporal pattern produced by both modulated signals. For greater electrode separations, the pitch was determined by the individual temporal patterns at each electrode. The implication of these data to coding of complex signals in sound processors is that pitch information derived from temporal information in the stimulus signals may be affected by phase differences between signals at neighbouring electrodes (McDermott, 2004). Specifically, out-of-phase rate information coded at neighbouring electrodes can result in reduced or inaccurate temporal pitch information and hence strategies should present temporal F0 information in phase across channels. Furthermore, those data suggest that coding of independent rate, or multiple concurrent rates, at nearby places is not possible. As a consequence, application of this research is limited to encoding of monophonic sounds in which a single rate is used to convey pitch, rather than to coding of chords and polyphonic sounds that comprise multiple rates.
More recently, Macherey and Carlyon (2010) examined the effect on rate-pitch of presenting temporal information on the same or adjacent electrodes (separated by 0.75 or 1.1 mm). They showed a greater independence between channels than was observed by McKay and McDermott (1996). The authors suggested several reasons for differences between studies the main ones being differences in procedure: McKay and McDermott employed a pitch scaling task which can be prone to non-sensory response bias whereas a pitch ranking procedure was employed by Macherey and Carlyon; and differences in devices: Nucleus 22 System straight array with electrode contacts that generally lie along the outer wall of the scala tympani compared to the HiFocus and Contour electrode arrays used in the more recent study for which electrode contacts are generally positioned closer to the auditory neurons. While Macherey and Carlyon concluded that the temporal codes conveyed by neighbouring electrodes are more independent than previously thought, some debate as to the underlying reasons for such outcomes still remain. The authors suggested that while spatial patterns of excitation to stimulation on neighbouring channels can be overlapping, it is possible that neurons conveying meaningful temporal information are located in spatially restricted regions that do not overlap much between channels. Alternatively, they suggested the responses to one or both channels spreads away, leading to “off-place listening”.

2.3.3 Amplitude modulation pitch

Like pulse rate, amplitude modulation (AM) rate of electrical pulse trains (see Fig. 2.4) can be used to elicit a percept of pitch (e.g., Shannon, 1992; McKay et al., 1994, 1995; Geurts and Wouters, 2001), although the salience of that percept can be weaker than that elicited by changes in pulse rate, particularly for shallow modulation depths. Furthermore, like pulse rate, a similar upper limit of approximately 300 Hz has been observed for modulation rate discrimination using sinusoidal amplitude-modulated (SAM) electrical pulse trains.

2.3.3.1 Effects of amplitude modulation depth on salience

Difference limens for amplitude-modulated pulse trains can be as low as those observed using pulse rate alone provided sufficiently deep modulation of the pulse train envelope is coded (e.g., Geurts and Wouters, 2001; McDermott and McKay, 1997). Considerable variation between subjects in terms of the modulation depths required for consistent discrimination of pitch has been observed (McKay et al., 1995; Geurts and Wouters,
On average, modulation depths ranging from 10 to 40% of the electrical dynamic range (EDR) are required to provide a salient percept of temporal pitch (i.e., performance similar to that observed for rate alone), although some subjects required depths of almost 100%. Studies examining pitch perception using clinical/commercial strategies have shown that they do not consistently code deep enough F0 modulation in the stimulus envelope to allow for provision of salient rate-pitch information (e.g., Green et al., 2004, 2005; Vandali et al., 2005) and thus the need for strategies that enhance the depth of F0 modulation cues to pitch.

2.3.3.2 Effects of stimulus duration on amplitude modulation pitch

Hall and Wood (1984) measured frequency discrimination of pure tones (500 and 2000 Hz) as a function of duration (from 5 to 200 ms) in a group of ten normal-hearing listeners and ten hearing-impaired listeners. Frequency difference limens for both groups improved monotonically with increasing stimulus duration up to approximately 100 ms, beyond which performance saturated. However, performance for the hearing-impaired listeners was poorer than that of the normal-hearing group. Similar findings were found by Freyman and Nelson (1986). The effect of stimulus duration on frequency discrimination was also measured using complex harmonic tones in normal-hearing listeners (White and Plack, 1998). For an F0 of 250 Hz, performance with stimuli in which harmonics were resolved improved with increasing duration up to some limit. So too did performance for stimuli in which harmonics were not resolved. However, little improvement for durations beyond 40 ms was observed for resolved harmonics stimuli, whereas for unresolved harmonic stimuli, improvements were seen for durations up to approximately 80 ms. In addition, it was observed that for unresolved harmonics, the duration to reach a maximum in performance increased with decreasing F0, consistent with a theory of temporal integration in which several-to-many repetition cycles are needed to determine frequency or pitch.

Lou et al., (2010) examined the effects of stimulus duration (50–400 ms) on amplitude modulation detection thresholds (AMDTs) and modulation frequency discrimination (AMFDTs) in eight adult CIs users. They observed little effect of duration on AMDTs although for some subjects, thresholds increased with the longer durations (200–400 ms). For AMFDTs at F0s of 50 and 100 Hz, performance improved with increasing duration from 50 to 100 ms, suggesting that a critical duration of 5-10 modulation cycles may be needed by CI users. For higher F0s (200 Hz), AMFDTs were
Chapter 2: Background and Relevance to Research

fairly constant across the stimulus durations tested. The trends in these data are similar to those observed for unresolved harmonic stimuli in acoustic hearing by White and Plack (1998) in which integration time is inversely proportional to modulation frequency.

The above outcomes demonstrate that the mechanism underlying temporal envelope pitch perception (i.e., derived from unresolved harmonics) in normal hearing responds more sluggishly over time as compared to that in which fine spectral and temporal structure is available in resolved harmonics. The mechanism underlying temporal pitch processing in electric hearing is thought to be similar (e.g., McKay and Carlyon, 1999) and both exhibit similar trends in frequency discrimination thresholds as a function of duration. Therefore in the application of rate-pitch coding, we should expect poor perception of pitch for short duration sounds (e.g., less than approximately 50 ms at 100 Hz). In addition, for sounds in which F0 changes rapidly, detailed pitch information may be smeared over time.

2.3.3.3 Effects of modulation shape on amplitude modulation pitch

Landsberger (2008) examined the effects of modulation shape on rate-pitch salience by measuring modulation rate difference limens (F0 DLs). Modulation shapes included sine, sawtooth, sharpened sawtooth (as per Green et al., 2004, 2005), and square modulation functions. F0 DLs were measured using a four-alternative-forced-choice (4AFC) procedure (e.g., Macmillan and Creelman, 1991) in which three intervals consisted of a 100 Hz modulated signal and the other a higher modulation rate. The subjects were asked to identify which interval was different from the others in any way other than loudness and so arguably was not a true measure of rate-pitch discrimination. F0 DLs were measured at three electrode places, apical, middle, and basal. DLs for the sine, sawtooth, and sharpened sawtooth modulation waveforms were similar. Poorer DLs were observed for square wave modulation. DLs (averaged across eight subjects) for sine, sawtooth, and sharpened sawtooth were approximately 6 to 9% for the three electrode places, although for the basal place using sawtooth modulation it was 18%. For square wave modulation, DLs were approximately 10% for apical and middle places, and increased to 25% at the basal place. While these outcomes demonstrated that rate-pitch DLs, which reflect the strength or salience of rate-pitch information, for sawtooth and sharpened sawtooth waveforms are similar to those for sine-wave modulation, little
Chapter 2: Background and Relevance to Research

is known about the pitch height (or relative pitch compared to unmodulated pulse trains) elicited by those stimuli.

2.3.3.4 Pitch-height of unmodulated and amplitude-modulated pulse trains

The pitch derived from amplitude modulation rate has been shown to be higher in general than that derived from unmodulated pulse-rate (e.g., McKay et al., 1995; Busby and Clark 1997; Swanson 2008). McKay et al., (1995) found that for shallow modulation depths, the pitch of a modulated stimulus was substantially higher than that of an unmodulated pulse train with a pulse rate equal to the modulation rate, but decreased towards that of the unmodulated pulse train as the depth of modulation was increased. It is presumed that this occurs due to greater phase locking (or entrainment) of neurons to the modulation rate, rather than the stimulation rate (or sub-multiples of the stimulation rate), as modulation depth is increased. Those data suggest that the pitch of AM stimuli is governed by some function of first order intervals in neural responses to the stimuli rather than the interval between peaks of the stimulus envelope. Those data are supported by outcomes of related studies which examined the pitch of irregular rate stimuli (e.g., Carlyon et al., 2002 and van Wieringen, et al., 2003) in which it was postulated that a weighted average of first order intervals in neural responses to the stimuli can be used to predict their pitch.

It is interesting to note that for the subjects and stimuli examined by McKay et al., (1995), the modulation depth required to match their pitch height to within one semitone (approximately 6%) was typically larger than the subject’s electrical dynamic range. This could have significant implications on the ability to accurately encode rate-pitch information using modulated pulse trains. Admittedly, the stimuli used in that study (i.e., low-moderate rate, 200-1000 PPS, bipolar electrode stimulation) were not typical of those used in existing devices (i.e., moderate-high rate, 500-5000 PPS, monopolar electrode stimulation) and it may be that closer pitch matches might be obtained for smaller modulation depths using higher stimulation rates and/or perhaps alternate modulation waveforms. For instance, the sharpened modulation functions (e.g., sawtooth waveform) examined by Green et al., (2004, 2005) and Landsberger (2008) may provide closer pitch matches to those of unmodulated pulse trains. However, those studies only reported data related to pitch strength and not pitch height and thus further research in this area is required.
Chapter 2: Background and Relevance to Research

2.3.3.5 Effects of pulse rate on amplitude modulation pitch

The stimulation pulse rate used to carry F0 modulation has also been shown to have a significant effect on pitch-height. McKay and Carlyon (1999) showed that for low carrier rates (i.e., below approximately 800 PPS) both modulation rate and carrier rate contributed to the rate-pitch percept. Furthermore, McKay et al., (1994) demonstrated that in order to adequately convey modulation rate information as a cue to pitch, the carrier rate must be at least four times the modulation frequency. Similar outcomes have been observed by other researchers (e.g., Wilson et al., 1997b). These data suggest that given an upper modulation rate-pitch limit of approximately 300 Hz, strategies need to employ stimulation rates of at least 1200 PPS per channel.

2.3.3.6 Effects of amplitude modulation depth on loudness

When coding pitch using amplitude-modulated pulse trains, the effect of modulation depth on loudness may be important. Using pulsatile electrical stimuli, McKay et al., (2003) showed that for low absolute current levels (less than about 200 µA for monopolar stimuli), loudness on a logarithmic scale was linearly proportional to log current. At higher absolute current levels, the relationship became progressively non-linear or exponential. This was also the case for sinusoidal amplitude-modulated (SAM) pulse trains. However for SAM stimuli (McKay and Henshall, 2010), higher level pulses within the modulation cycle contributed more to the overall loudness than those in the same modulation pattern presented at lower levels. Hence the relationship between loudness and stimulation level for modulated stimuli is also dependent on modulation depth. For experimental strategies which increase the modulation depth coded in channels, subsequent changes in channel loudness may thus need to be compensated for so that overall loudness remains consistent with that produced by clinical strategies. However, because modulation rates examined by McKay and Henshall (2010) were higher than those typically associated with temporal envelope cues to pitch, further work in this area is needed using modulation rates more relevant to the range of F0s that can be encoded via rate-pitch. In addition, if functions other than sinusoidal modulation of current are to be used in experimental strategies (e.g., sharper functions such as sawtooth modulation), then the effects of those modulation shapes on loudness also need investigation.
2.3.4 Relevance to research

The implication of the above outcomes to coding of pitch in CIs is that while place coding may be sufficient to convey the pitch of some sounds such as pure tones and band-pass filtered noise, resolution of fine spectral detail, such as F0 harmonics in complex sounds, does not seem possible with present CI devices. The main reasons for this stem from the limited spectral resolution and selectivity provided by place coding due to the broad pattern of excitation to electrical stimulation (§2.3.1.2), the variability in position of electrodes relative to auditory neurons, and poor auditory survival across subjects (§2.3.1.1). Place coding is thus limited to coding of spectral envelope information in general, at least for low F0s. For higher F0s where changes in place with F0 can be large enough to be discriminated, place coding may provide some benefits. However, because changes in place are not always consistent with changes in F0, the mechanism has limited applicability for real-world sounds in which F0 and the spectral envelope can vary independently. For these reasons the present research was directed towards utilising rate cues to code pitch.

For low F0 frequencies, up to around 300 Hz, changes in electrical pulse rate have been shown to elicit similar ratiometric changes in pitch to those observed for F0 in normal hearing (§2.3.2.3). Average pulse-rate discrimination thresholds for moderate-to-excellent performing CI recipients are close to those needed to follow the smallest melodic intervals in music and to discriminate voicing frequency in speech (§2.3.2.2). However, the upper rate limit in which pulse-rate discrimination deteriorates (§2.3.2), the weaker salience (§2.3.2.2), and the sluggish response (§2.3.3.2) of the percept elicited by rate compared to that provided by fine spectral and temporal information in normal hearing, together with the inability to code rate information independently at nearby spatial locations (§2.3.2.8), all restrict the application of this research to coding of the typical voicing frequency range (up to approximately 300 Hz), to encoding of F0 for monophonic pitch, and to encoding of pitch in sounds where F0 does not change too rapidly.

Because use of low pulse rates compared to the higher pulse-rates employed in present clinical strategies can degrade perception of spectral envelope and low-frequency temporal envelope information (which will be discussed later in sec 2.6.22.1), it is preferable to use higher pulse rates for which F0 information is encoded in the temporal envelope of the stimulus signal. This is achieved via amplitude-modulation of a
high-rate pulse train (§2.3.3). Changes in amplitude-modulation rate have been shown to produce similar changes in pitch to those observed for pulse-rate provided that sufficiently deep modulation is employed (§2.3.3.1) and that the carrier pulse rate is sufficiently high (§2.3.3.5). However, it is unclear whether and for what conditions the pitch-heights elicited by amplitude-modulated pulse trains are similar to that of unmodulated, low-rate, pulse trains (§2.3.3.4). It is hypothesised that use of amplitude modulation characterised by a sharp onset and rapid exponential decay, so that each F0 interval is primarily imparted by a single electrical pulse should provide a salient (§2.3.3.3) and accurate (§2.3.3.4) representation of F0 in the auditory system from which pitch can be derived. This topic was examined by the present research in chapters 3 and 5. In addition, the effect of this type of modulation on loudness was also examined in those chapters. The latter was done because the application of deep/sharp amplitude modulation in the stimulus envelope was expected to alter loudness (§2.3.3.6) and hence potentially affect coding of spectral envelope and low-frequency temporal envelope information when applied in an experimental sound coding strategy. Furthermore, because place and level of stimulation, which vary with different signals, can affect rate-pitch (§2.3.2.5; §2.3.2.6 §2.3.2.7), their influence on rate-pitch was also examined in chapter 3.

2.4 Music perception using clinical processing strategies

Musical sounds are comprised of independent dimensions or attributes that can be extracted by specialised processing in the brain. These attributes are pitch (which governs melody and harmony), timbre (tone colour or quality), rhythm (and its associated concepts tempo, meter and articulation), dynamics (volume and slow changes in loudness), and texture (the combination of melodic, rhythmic, and harmonic elements in a composition). The present research is concerned mainly with the first two of those attributes.

2.4.1 Pitch perception

While variations in pulse-rate, modulation rate, and/or place of stimulation, in a psychophysical context, can be used convey changes in pitch, studies examining pitch perception when sounds are processed through subject’s clinical processors have demonstrated poorer and more variable performance (Gfeller and Lansing, 1991; Pijl, 1997b; Leal et al., 2003; Vandali et al., 2005; Sucher and McDermott, 2007; Gfeller et
Chapter 2: Background and Relevance to Research

al. 2002a, 2005, 2007; Galvin et al., 2007; Wei et al., 2007; Swanson, 2008; Lee and Lee, 2010; Wang et al., 2011).

2.4.1.1 Pitch interval perception

Pijl (1997b) conducted a pitch interval labelling experiment, similar to that described earlier (§2.3.2.3), with two CI and four NH adult subjects. For the CI users, the electrical stimuli consisted of pairs of low-rate electrical pulse trains (ranging from 54 to 266 PPS) presented sequentially at a single electrode site. For the NH listeners, acoustically equivalent pairs of click trains were used. Both the CI subjects and the NH subjects labelled rate intervals constructed from those stimuli (5th, 4th, and minor 3rd) with a high degree of accuracy. However, when the experiment was repeated with the same CI subjects listening to acoustic stimuli produced by a piano and processed through their clinical sound processor SPEAK strategy (Skinner et al., 1991; Seligman and McDermott, 1995) they were unable to consistently judge whether the interval was in tune or not. That results was attributed to poor coding of F0 information by the SPEAK strategy and is not surprising given that F0 is not explicitly coded using pulse-rate or place of stimulation by the strategy, and that the strategy employs a low rate of stimulation resulting in poor sampling of any temporal F0 information in the stimulus envelope.

2.4.1.2 Melodic pitch perception

Melodic pitch perception was examined by Gfeller and Lansing (1991) in a group of eighteen adult CI users (eight of whom used a 4-channel vocoder strategy with the Ineraid implant system and the other ten used the F0/F1/F2 speech-feature extraction strategy with the Nucleus 22 implant system). Pitch discrimination was examined using the Primary Measures of Music Audiation (PMMA) test (Gordon, 1979) which consists of 40 electronically produced pairs of melodic patterns. Each pair contained from two to five notes ranging in F0 from C4 (260 Hz) to F5 (694 Hz). Each pair had identical temporal patterns but one or more of the notes within the pattern could differ in F0. The pairs were presented in random order and the subject was asked to nominate whether the two patterns were the same or different. The mean subject score was 78% correct and scores were similar between both groups of implant users. The authors however noted that those scores were lower than those obtained by normal-hearing listeners on the same test. In addition, given that the range of F0s examined were at and beyond the upper
Chapter 2: Background and Relevance to Research

limit of rate-pitch discrimination in CIs then it is possible that subjects may have utilised changes in place of stimulation to judge the pitch of those synthetic stimuli.

Leal et al., (2003) examined the ability of twenty-nine adult CI users, twenty of whom used the ACE strategy and the other nine used the SPEAK strategy, to discriminate and identify pitch differences between a pair of short musical pieces. Each pair of pitch sequences had identical temporal patterns but their constituent notes differed in F0 across pairs by either 6, 8, 12, or 24 semitones, or by 4 or 5 semitones for three notes only within the sequence. In the discrimination test, subjects were asked whether the pair was the “same” or “different”. In the identification test, they were asked whether the pitch became “higher” or “lower” across the pair and whether this change occurred at the “beginning”, “middle”, or “end” of the sequence. For the majority of subjects, results of the pitch discrimination test were very good with an average score across all F0 intervals of 91.7% (close to 100% for largest F0 differences of 12 and 24 semitones and 86% for 6 semitone intervals). For the pitch identification test, mean scores for a subset (three quarters) of the subjects was 79.2% (close to 100% for 24 semitone intervals and 62% for 6 semitone intervals). While those data demonstrated reasonable outcomes (albeit for large F0 intervals) compared to other related studies, it is difficult to gauge their importance given that no normative data for normal-hearing listeners on these tests are available and that no details were provided about how the stimuli were generated and what range of F0s they spanned. For example, if synthetic stimuli were used and their F0 was high enough so that large and consistent changes in the mean place of stimulation across the pairs were produced, then we might expect subjects to perform well based on changes in place-pitch. However, the same is not likely to be true for sounds produced by real musical instruments in which variations in resonant frequency with F0 are not necessarily consistent within and across instruments.

2.4.1.3 Pitch ranking of sung-vowels

A number of researchers have examined pitch perception using more natural complex harmonic tones, such as sung vowels (e.g. Vandali et al., 2005; Sucher and McDermott, 2007). Vandali and colleagues examined pitch ranking of sung-vowels with a group of five adult CI users of the SPEAK or ACE strategy. The sung vowels used in that test were representative of a natural sung voice embodying slight variations in spectral-timbre (and hence place coding). Two vowels, /a/ and /i/, were sung by a male and a
female choir singer. For each vowel, four notes were produced, starting from G2 and increasing progressively in half-octave (six semitone) steps to C#3, G3, and C#4 for the male, and C4, F#4, C5, and F#5 for the female. F0s for the four notes were 98, 139, 196, and 277 Hz for the male and 262, 370, 523, and 740 Hz for the female. The four tokens for each vowel were combined into pairs that differed in F0 by half an octave. Using a two-interval, two-alternative, forced-choice procedure, the stimulus pairs were presented in a randomised order and subjects were asked to nominate which of the two sequentially presented sounds was higher in pitch. Instructions were given to ignore any variations in loudness and no feedback was provided about the correctness of responses. The group average percent correct scores for the male and female stimuli were 69.1% and 73.6% respectively. Using a two-tailed Binomial distribution (N = 32) to determine whether the proportion of correctly ranked responses was significantly different from a chance score of 50%, results were only marginally above the threshold score (for \( p = 0.05 \)) of approximately 65% for that test. While results for the male stimuli, which spanned a range of F0s below 300 Hz, could be partly attributed to perception of temporal F0 cues, the same could not be said for the female stimuli because their F0s were beyond the range coded by temporal information in the clinical strategies. It is thus likely, that subjects attended to changes in place of stimulation to judge pitch differences, which for some pairs were consistent with changes in F0 but were not for others. It is also likely that even though F0 temporal cues were available for the male stimuli, changes in place across stimuli affected the subject’s judgment of pitch.

Pitch ranking of sung-vowels was also examined in a group of ten normal-hearing listeners and compared to performance of eight CI recipients using their clinical SPEAK strategy (Sucher and McDermott, 2007). The same sung vowels as described in Vandali et al. (2005) were used in the pitch ranking tests but F0 intervals of one semitone (ranging from C#3: 139 Hz, to G3: 196 Hz for the male singer and from F#4: 370 Hz, to C5: 523 Hz for the female singer) as well as six semitones (as described earlier) were examined. The normal-hearing listeners obtained significantly higher percent correct scores than the CI recipients for both sets of F0 intervals (NH: 81.2% and 89.0% for one and six semitone intervals respectively compared to CI: 49.0% and 60.2% respectively). For the normal-hearing listeners, prior musical experience was shown to be correlated with higher pitch ranking scores. However, even the musically inexperienced NH listeners obtained significantly higher scores than the CI subjects for both sets of F0
Chapter 2: Background and Relevance to Research

intervals suggesting that the pitch information available to those CI users was less accurate than that available in normal hearing.

2.4.2 Melody recognition

2.4.2.1 Perception by cochlear implant and normal-hearing listeners

In a series of studies, Gfeller and colleagues examined familiar melody/song recognition by cochlear implant and normal-hearing listeners and compared their performance to other perceptual measures including pitch discrimination of pure tones or complex harmonic tones, and speech recognition. For CI recipients, melody recognition was also compared to other factors including age at time of testing, duration of profound deafness and musical experience post implantation. Recognition of twelve familiar melodies and twelve foils generated using the durational values and pitches of each familiar melody, but in a new sequential order, was examined by Gfeller et al. (2002a). A group of eighteen normal-hearing and forty-nine cochlear implant adults who used either the SPEAK, CIS or ACE processing strategy participated in the study. Each familiar melody and its foil was presented without lyrics as the melody only and as the melody plus harmony. The melodies were presented in the standardised key of C major (with an F0 range of G3: 196 Hz to C5: 523 Hz) and were synthesised using a standard MIDI acoustic piano. The mean percentage of melodies recognised by the NH listeners was approximately 55% (or 83% if adjusted by eliminating those melodies not known by individuals) which was significantly higher than that for the CI recipients who scored approximately 13% (or 19% if adjusted). No significant differences between recognition of melody only versus melody plus harmony was observed nor between each subgroup of CI device or processing strategy. Pitch ranking of complex harmonic tones, with F0s ranging from D2 (73 Hz) to C#5 (553 Hz), synthesised using a standard MIDI acoustic grand piano was also examined in a subset of those subjects (eight NH and forty-six CIs). An adaptive two-interval-forced-choice (2AFC) test was used to determine F0 discrimination thresholds. The mean F0 discrimination threshold for the NH listeners was 1.13 (range 1-2) semitones compared to 7.56 (range of 1-24) semitones for the CI recipients. A significant but moderate correlation (r = -0.50) between familiar melody recognition and complex harmonic tone discrimination was observed. For the CI users, significant negative correlations were also observed between melody recognition and
both age at time of testing and length of profound deafness, and a positive correlation was observed between melody recognition and speech recognition scores.

A similar study to the one described above was conducted with a larger group of subjects (Gfeller et al. 2005). Open-set recognition of selected musical excerpts (without lyrics) was examined in a group of thirty NH adults and seventy-nine CI recipients divided into groups according to implant type (Nucleus, Clarion, or Ineraid) and processing strategy (SPEAK, CIS, or ACE). The musical excerpts were taken from three musical genres: classical, country, and pop. For all three genres, NH listeners were significantly more accurate at identifying melodies than the CI users. There was no significant difference in recognition of musical genre by the NH listeners but a significant reduction in recognition of classical, compared to country or pop items, was observed for the CI recipients. There were no significant differences amongst CI users due to implant type or processing strategy. For CI users, moderate to weak significant correlations between melody recognition and age at time of testing, speech perception, and musical experience post implantation were observed.

2.4.2.2 Perception by cochlear implantees with residual hearing

Melody recognition in CI subjects with some residual hearing was examined by Gfeller et al. (2006, 2007). In the first of those studies, melody recognition using real-world songs (country and pop presented with or without lyrics) by four CI recipients with some residual hearing (who used a hybrid cochlear implant plus hearing aid device) was similar to that of a group of seventeen NH listeners, and significantly better than that of a group of thirty-nine CI alone recipients. The latter result was however only the case for songs presented without lyrics. A follow-up study (Gfeller et al., 2007) was conducted with a larger cohort of subjects comprising, twenty-one NH listeners, one-hundred and one standard (long-electrode array) CI alone users, and thirteen (short electrode) CI recipients that received acoustic plus electric stimulation (A+E). The authors however did not directly compare melody recognition across the subject groups. Instead, pure tone pitch discrimination and pitch ranking for base frequencies of 131 to 831 Hz, and interval sizes of one to four semitones, were measured in each subject. Pitch discrimination and ranking by the CI alone group was less accurate than that of the NH and A+E groups. For stimuli in the higher frequency range where the A+E subjects had little residual acoustic hearing, their performance deteriorated towards that of the CI alone group. For the CI users in both groups, pitch ranking of the pure tones was
significantly correlated with melody recognition and speech perception in noise and demonstrated the practical benefits of preserving low frequency hearing.

Other studies examining musical perception skills of cochlear implantees, normal-hearing listeners, and hearing-impaired listeners using hearing aid(s) have all demonstrated similar results to those of Gfeller and colleagues (e.g., Leal et al., 2003; Kong et al., 2004; Looi et al., 2008; and Won et al., 2010). For instance, Looi et al. (2008) examined rhythm and pitch discrimination as well as instrument and melody recognition in a group of fifteen CI and fifteen hearing aid (HA) users that had similar levels of hearing impairment. Results of rhythm discrimination and instrument recognition tests were not significantly different between the two groups of subjects. However, both pitch ranking and melody recognition was significantly poorer for the CI as compared to HA users with mean melody recognition scores of 52% and 91% respectively. Those results further highlight the importance of acoustic information to processing of pitch and melody.

2.4.2.3 Importance of spectral and temporal information to pitch and melody
The relative importance of spectral and temporal information coded by clinical strategies to perception of music attributes and melody was examined by Won et al. (2010). In a group of forty-two CI users, discrimination of pitch and timbre, and recognition of melody was measured and correlated with measures of spectral and temporal processing using spectral-ripple (Won et al., 2007; Henry et al., 2005) and Schroeder-phase discrimination (Drennan et al., 2008) tests respectively. Spectral ripple discrimination was shown to correlate with pitch, timbre and melody perception but this was not the case for Schroeder-phase discrimination. However, the lowest F0 examined in the music perception tests was around middle C (262 Hz) where coding of temporal cues to F0 is poor in clinical strategies and where rate-pitch discrimination in CIs begins to deteriorate rapidly. This would explain the lack of any correlation between temporal processing and the results of the music tests.

2.4.3 Timbre perception
In music, timbre or tone colour/quality is an attribute of an auditory sensation which characterises differences, other than pitch or loudness, in sounds produced by different musical instruments. Perception of timbre is usually measured through recognition and
2.4.3.1 Perception by cochlear implant and normal-hearing listeners

Gfeller and colleagues (2002b) examined the effects of instrument type and frequency range on timbre recognition and appraisal in normal-hearing listeners and cochlear implant users. Recognition and appraisal of eight different musical instruments representing three F0 ranges (low: 131-262 Hz, medium: 262-534 Hz, and high: 534-1068 Hz) and four instrument families (brass, woodwind, pitched percussion, and string) was examined. Instruments included the trumpet (high), trombone (low), flute (high), clarinet (medium), saxophone (low), piano (high and low), violin (high), and cello (low).

In the recognition test, subjects were required to identify the instrument from a set of sixteen possible instruments. The NH listeners achieved a mean score of around 91% correct which was significantly higher than that of the CI recipients whose scores were around 47%. In the appraisal task, ratings of overall pleasantness for the NH listeners were significantly higher than those for the CI users, although significant differences within and across groups were also seen across the different F0 ranges. For the CI users, poorer and more scattered ratings for sounds in the high F0 range (i.e., string instruments) were observed. The authors suggested that those sounds may elicit a "distorted", "harsh", or "shrill" sound rather than a more "brilliant" sound as reported by the NH listeners.

2.4.3.2 Perception by cochlear implantees with residual hearing

Instrument recognition was examined by Looi et al., (2008) in a group of fifteen CI and fifteen hearing aid (HA) users that had similar levels of hearing impairment. Three tests were performed that measured recognition of single instruments, solo instruments with a background accompaniment, and music ensembles each playing as a cohesive, unified group. In each test, twelve different instruments or music ensembles were presented four times in randomised order and subjects were required to select the instrument or ensemble playing from a list of twelve matching pictures. The mean percent correct scores of the three tests for the CI recipients were 61, 45, and 43% respectively and were slightly lower than those for the HA group who scored 69, 52, and 47% respectively. There was no significant difference between each group’s scores, although scores for recognition of single instruments were significantly higher than those of the other tests.
Chapter 2: Background and Relevance to Research

for both groups of listeners. The authors indicated that the comparative score for these tests by NH listeners was around 95% and suggested that poorer results obtained by both hearing-impaired groups was indicative of insufficient transmission of spectral and/or temporal envelope information to enable listeners to accurately identify timbre.

2.4.3.3 Effects of timbre on pitch

Although perception of timbre is not a main goal of the present research, its importance cannot be overlooked given its influence on perception of pitch in CI hearing (e.g., Sucher and McDermott, 2007; Galvin et al., 2008) and normal hearing (e.g., Singh and Hirsh, 1992; Warrier and Zatorre, 2002; Russo and Thompson, 2005; Pitt, 1994). For instance, Pitt (1994) found that non-musically trained normal-hearing listeners often confused changes in timbre as a change in both pitch and timbre, and vice-versa, although that was rarely the case for musically trained listeners. Galvin and colleagues (2008) observed that CI recipients’ identification of melodic contour was affected by instrument timbre particularly for those subjects with the least musical experience. Mean subject performance was best with the organ and poorest with the piano. Although mean scores were not significantly different across instruments, individual subject scores were significant in five of the eight subjects. While those data show that changes in timbre can influence listeners’ judgement of pitch, the data showing less effect in musically experienced listeners suggests that music training may help less experienced listeners to perceive pitch in music and sounds in which timbre can vary.

2.4.4 Summary

When using clinical sound processors and strategies, cochlear implant recipients generally exhibit poorer pitch (§2.4.1.1), pitch contour (§2.4.1.2) and melody discrimination (§2.4.2) compared to normal-hearing listeners, although the inclusion of lyrics and rhythmic cues can aid melody recognition. Perception of pitch and melody by CI recipients can be improved by addition of acoustic hearing using a contralateral hearing aid or hybrid acoustic plus electric device (§2.4.2.2). Timbre perception and instrument recognition (§2.4.3) by CI recipients is poor compared to NH although not as bad as that of pitch perception. Moderate to weak correlations have been found between melody discrimination abilities and pitch discrimination, speech recognition in noise, musical experience, duration of profound deafness, and age at time of testing (§2.4.2.1).
2.5 Tonal language perception using clinical systems

In tonal languages, voice pitch variations within the same phonemic segment produce a change in lexical meaning (e.g., Gandour, 1981). Tonal languages are spoken in almost every continent of the world and in the Asia Pacific region there are an estimated 1.3 billion people who speak a tonally based language, including those of China, Hong Kong, Taiwan, Vietnam, Thailand and Laos. Mandarin and Cantonese are by far the most commonly spoken tonal language amongst people throughout the world. Mandarin consists of four lexical tone contours: Tone 1 (high flat); Tone 2 (middle low rising); Tone 3 (dipping and rising); and Tone 4 (high falling), see Fig. 2.6. Cantonese has six major tone contours: High-Level (HL); High-Rising (HR); Mid-Level (ML); Low-Rising (LR); Low-Level (LL); and Low-Falling (LF), see Fig. 2.7. There is also a high-falling tone which is not as commonly used in the dialect. The primary acoustic cues to tone recognition are F0 height and contour (e.g., Liang, 1963) but secondary cues, such as duration, amplitude contour, and spectral envelope, are also utilised particularly when F0 information is corrupted for example by competing noise or when F0 is poorly perceived (as is the case for CI recipients).

Figure 2.6. Mandarin lexical tone contours taken from Wei et al., (2004) for the same consonant-vowel syllable /ma/ meaning “mother”, “linen”, “horse” and “curse” for each tone respectively.
Figure 2.7. The six main Cantonese lexical tone contours taken from Francis et al., (2008) for an adult male native speaker.

2.5.1 Mandarin

Recognition of Chinese-Mandarin tone by CI recipients has been examined (e.g., Huang et al., 1996; Fu et al., 2004; Wei et al., 2004, 2007; Wang et al., 2011; Xu and Zhou, 2011) and in general, performance has been shown to be significantly worse than that of NH listeners, although considerable variation amongst subjects has been observed which is typical in CI hearing. Wei et al., (2004) examined recognition of Mandarin tone in a group of five adult CI recipients of the Nucleus 22 system who used the SPEAK strategy. Twenty-five monosyllabic words each having four tonal patterns were produced by a male speaker. Each token was presented five times in randomised order in a 4AFC procedure and subjects were asked to identify the tone. A total of nine experimental speech processor maps were tested which included four single channel maps (with each channel at a different electrode site) and five maps in which 4, 7, 10, 14, or 20 electrodes were used. Mean percent correct results of the tone recognition tests were near chance (25%) for all four single channel maps and increased monotonically to plateau at around 70% for 7- to 20-electrode maps. The scores for the 7- to 20-electrode maps were significantly higher than those of the other maps. NH listeners who
performed the same test listening to acoustic CI simulations of each experimental map (Kong et al., 2003) obtained substantially higher scores with fewer numbers of channels (around 70% for single channel and increasing to around 90% for 7 and 14 channel maps).

2.5.1.1 Tone recognition in noise

Lexical tone recognition in quiet and noise was further examined by Wei et al., (2007) in a group of seventeen CI recipients using the ACE strategy and five NH listeners. The same tone recognition test described in their earlier study was used. Speech shaped noise was used at signal-to-noise ratios (SNRs) of +10 to -10 dB in 5 dB steps. Average tone recognition scores of around 80% (for postlingually deaf subjects) were obtained in quiet which dropped to around 55% at +10 dB SNR and to chance level performance (around 25%) at -5 dB SNR. Those results were substantially poorer than those of NH listeners who maintained nearly perfect performance across SNRs except at -10dB SNR in which they scored around 74% correct.

2.5.1.2 Tone recognition correlations

Recognition of Mandarin lexical tones and pitch discrimination was measured with nineteen CI recipients using their everyday processor (SPEAK, ACE, or HiResolution) and ten NH listeners by Wang et al. (2011). Pitch discrimination was measured in a melodic context using a same/different interval discrimination task. The melody was presented twice, either as the same melody, or with the F0 of one note in one melody altered. Subjects were required to indicate whether the two melodies were the same or different. Two simple melodies were used, Twinkle Twinkle Little Star and Happy Birthday, which spanned F0 ranges of 131 to 220 Hz and 524 to 699 Hz respectively. The notes in each melody were synthesised complex harmonic tones consisting of F0 and its first three harmonics. The mean pitch discrimination threshold for sixteen of the nineteen CI recipients that could perform the test was 5.7 semitones (6.8 and 4.5 semitones for the two melodies respectively) and ranged from 0.8 to 19.6 semitones across the subjects. The NH listeners performed significantly better with thresholds ranging from 0.24 to 0.91 semitones and a mean of 0.44 semitones. Mandarin tone recognition was measured using ten monosyllabic syllables spoken in each of the four tones by a male and female speaker. The F0 range of the tones was 80-250 Hz and 150-350 Hz for the two speakers respectively. The duration of the four tones were balanced.
within each syllable. Each token was presented twice in randomised order using a 4AFC procedure. For the nineteen CI recipients, mean tone recognition scores of 58.3% with a range of 12.5-86.8% were reported. In contrast, the NH listeners performed nearly perfectly with a mean score of 97.3% which was significantly higher than the CI group. Tone 2 (rising) was the most commonly confused token, often confused with Tone 1 (flat) or Tone 4 (falling). Tone recognition scores for the CI users were negatively correlated with age and duration of deafness and positively correlated with duration of hearing aid use for those subjects that wore a hearing aid prior to implantation.

Wang and colleagues (2011) also found that tone recognition was negatively correlated with pitch interval discrimination. That correlation was similar between the two melodies examined despite their different F0 ranges. The author suggested that subjects may have been able to utilise temporal modulation cues to pitch for the low F0 range and place cues for the higher F0 range. They also indicated that subjects may have used cues other than F0, such as the amplitude contour which is known to be correlated with tone (e.g., Whalen and Xu, 1992). Nevertheless, the strong correlation observed between musical and lexical pitch perception in CI hearing suggests that they may share a similar mechanism in auditory processing, although that finding is not supported by a study with on Mandarin-speaking normal-hearing musicians which showed no correlation between musical pitch and lexical tone perception (Lee and Lee, 2010).

2.5.1.3 Tone recognition in CI children

Xu and Zhou (2011) presented the results of a study in which Mandarin lexical tone recognition in one-hundred and seven prelingually-deaf CI children aged from 2.4 to 16.2 years old was examined. The test involved identifying one of the four Mandarin tones from six tone contrasts using a 2AFC procedure. In each trial, a tone was played acoustically while pictures of the two tones within the tone contrast being tested were displayed. The child was required to select the picture corresponding to the meaning of the word they had heard. Chinese words familiar to the age group of the children were used. Performance of individuals varied from near perfect scores to chance with a group average of 67% correct which was significantly lower than the near perfect results obtained by a group of one-hundred and twelve NH children. The CI children performed worst for contrasts containing Tone 1, which was overall the best recognised tone.
2.5.2 Cantonese

2.5.2.1 Tone recognition

The perception of Cantonese lexical tones was examined by Ciocca et al., (2002) in cochlear implantees with early-onset deafness. Seventeen children aged between 4.5 and 8.9 years old who had been using the Nucleus CI device for 0.9-3.4 years took part in the study. Six subjects used the SPEAK strategy and the other eleven used the ACE strategy. Cantonese lexical tone recognition was measured using a 2AFC picture identification procedure (similar to that described in §2.5.1.3) in which subjects identified one of six contrastive tones produced in the monosyllable /ji/. The six target tones were grouped into eight tonal contrasts: HL-ML, HL-LL, ML-LL, HR-LR, LR-LL, LF-LR, LF-LL, and HL-HR. The first three contrasts were used to examine effects between the three pitch levels (high, mid, and low) on tone perception. The next four contrasts were used to examine sensitivity to F0 differences at the end of the tone whereas the final contrast focused on F0 differences at the start of the tone. Mean percent correct scores for each tone contrast ranged from 50 to 61%. Group performance was only significantly above chance for three contrasts, HL-ML, HL-LL, and HL-HR, and only a few children scored above chance according to a binomial test (> 75%). Mean scores for the HL-ML tonal contrasts were significantly higher than those for the ML-LL and HR-LR contrasts but not for any of the other pairwise comparisons. The authors suggested that listeners were more accurate at identifying tones when the alternatives were separated by large F0 differences.

2.5.2.2 Word recognition

A longitudinal study examining spoken word recognition by Chinese children implanted with a cochlear implant at different ages was conducted by Lee et al., (2002). Sixty-four prelingually deaf children implanted at ages ranging from 1.1 to 14.1 years old participated in the study. The children were divided in three groups according to age at time of implantation: (1) < 3 years; (2) 3 to 6 years; and (3) > 6 years. Open-set Cantonese word recognition was measured before implantation and six times thereafter over a period of five years. Continuous improvement over the five year period in word recognition was observed in all children. Children implanted earlier than 3 years of age improved at a slower rate in their first year of implant use as compared to children implanted at an older age. However by two years of use the younger children’s
performance exceeded that of the older children and that trend was maintained over the remainder of the five year period. Those data were consistent with data for English speaking children (e.g., Geers, 2004) in which better outcomes are obtained with early implantation and longer implant experience.

2.5.3 Effects of stimulation rate

Several researchers have examined the effects of stimulation pulse rate on perception of lexical tone using clinical coding strategies such as CIS, SPEAK and ACE that encode F0-related cues using amplitude modulation in the stimulus envelope (Barry et al., 2002; Au, 2003; Fu et al., 2004). While no overall significant effect of stimulation rate within strategy was observed in those studies, individual variability across subjects was present, and a between strategy effect was reported by Fu and colleagues. They found that Mandarin tone recognition scores for the SPEAK strategy using a stimulation rate of 250 PPS/CH were significantly lower than those for ACE (at rates of 900, 1200, and 1800 PPS/CH) and CIS (at rates of 1200, 1800, 2400, and 3600 PPS/CH). The authors attributed that result mainly to better temporal sampling of F0 information by the higher rate ACE and CIS strategies. Au (2003) reported a trend of improvement in Cantonese lexical tone recognition scores across rates of 400, 800 and 1800 PPS/CH for a group of 11 adult Cantonese speaking CI recipients. The author suggested “that the maximum rate of 1800 PPS/CH could be an optimal stimulation rate and informed choice of parameter for the benefit of Cantonese-speaking CI users in lexical tone perception.”

2.5.4 Summary

Perception of Mandarin lexical tones by native adult Chinese CI recipients using their clinical processing system is generally poorer, around 70-80% correct, than NH listeners who obtain near perfect perception (§2.5.1). Their performance in noise drops markedly for SNRs of around +10 dB and by -5 dB scores are at chance level (§2.5.1.1). Prelingually-deaf children implanted with a CI also perform poorer than NH children with average scores of around 70% and 50-60% correct for Mandarin (§2.5.1.3) and Cantonese (§2.5.2.1) lexical tone contrast identification respectively. Factors influencing lexical tone recognition include: age and duration of deafness (younger and shorter period of deafness is better, §2.5.1.2); age at time of implantation and CI experience (early implantation and longer experience is better, §2.5.1.3; §2.5.2.2) and pitch interval discrimination (lower thresholds are better, §2.5.1.2). Considerable variability across CI
subjects is observed for lexical tone recognition as a function of stimulation rate which is consistent with performance in non-tonal languages speaking subjects (§2.5.3). Some results do however show that low stimulation rates of 250 PPS/CH do not adequately sample temporal F0 information and suggest that higher stimulation rates might therefore be optimal as a clinical default rather than a lower rate. The minimum numbers of channels/electrodes in clinical CI strategies for which lexical tone recognition plateaus is around seven (§2.5.1) which is similar to that seen for speech perception in non-tonal language speaking CI users (e.g., Friesen et al., 2001).

2.6 Pitch coding strategies

2.6.1 Place-pitch coding strategies

Despite the limited place-coding resolution and sensitivity available in electrical hearing, researchers have examined the possibility of using electrode place to code more detailed information about fundamental frequency (F0) and hence pitch.

2.6.1.1 Finer spectral resolution

Geurts and Wouters (2004) utilised a band-pass filter bank with sufficient spectral resolution in the F0 frequency region to resolve the frequency of the fundamental in any two adjacent filters. The envelope signals from these filters were allocated to electrodes of an eight channel CIS strategy. The range of F0s examined in this study was 110 to 263 Hz and hence fell within the range in which temporal cues to F0 could also be utilised. In the absence of temporal F0 cues (by filtering the envelope signals using a 20 Hz low-pass filter), the experimental filter bank provided lower F0 discrimination thresholds for synthetic vowel stimuli compared to the conventional CIS filter bank approach. However, when temporal cues to F0 were reintroduced by removing the envelope signal low-pass filter, differences in detection thresholds between filter banks were reduced indicating that the temporal cues provided substantially similar levels of information about F0, at least for low F0s. Similar outcomes were obtained by applying this type of filter bank to the ACE coding strategy in which up to 22 electrodes are available for stimulation using the Nucleus CI24 cochlear implant (Lanaeu et al., 2006). Those outcomes suggest that place coding can provide some benefits to pitch perception, especially for CI users who have good place-pitch sensitivity and poor rate-pitch discrimination. However in general, provision of temporal cues to F0 would appear to
make up for lack of finer place cues, at least for F0s within the range that can be discriminated using rate. In addition, these results are still substantially worse as compared to normal-hearing listeners.

### 2.6.1.2 Musical interval resolution

In a study by Kasturi and Loizou (2007), the effect of F0 place coding on melody recognition by CI recipients was examined. Melody recognition was measured using subject’s standard CIS filter bank and an experimental one in which bands were spaced by semitone frequency intervals. Results were significantly higher for a 12 channel filter bank with semitone spacing compared to the spacing used in subjects’ standard processor. However, unlike the earlier studies described above, the F0 of the material tested in this study was beyond the range that could be perceived using temporal cues to pitch. The mean F0 of the melodies tested was 440 Hz (A4) plus or minus one semitone with a range of 12 semitones. Thus subjects most likely could only utilise place cues to judge pitch, which seems like a reasonable approach for coding of F0s beyond approximately 300 or 400 Hz.

A similar study was conducted by Omran et al., (2010) in which melody contour and instrument recognition tests were conducted with two-different semitone-interval based filter banks and compared to the standard filter bank arrangement used in ACE. The first experimental filter bank (Smt-LF) restricted the frequency range coded between 130-1502 Hz, the second (Smt-MF) between 440-5009 Hz, and the standard filter bank frequency range was 125-8000 Hz. The two experimental filter banks consisted of 43 semitone spaced channels allocated to the 43 dual electrode channels available in the Nucleus Freedom CI24RE (Busby and Plant, 2005). A total of eight adult CI recipients were tested. Melodic contour identification was examined using a modified version of the MCI test (§2.7.3.1; Galvin et al., 2007) in which only five of the nine melodic contours were included. An F0 range of around 150 to 350 Hz was tested with interval sizes of 1, 2 or 3 semitones between notes within each pattern. Mean results showed no effect of filter bank mapping on MCI scores. Instrument recognition was examined using a familiar piece of music played by eight different instruments divided into four families. A significant reduction in overall instrument identification scores was observed using the Smt-LF map compared to the standard ACE map. Subjects reported that sounds elicited by the Smt-LF map were much higher in pitch than normal and some disliked it. No other significant differences were found.
2.6.1.3 *Limitations of place-pitch coding strategies*

There are a number of drawbacks associated with F0 place coding. First, the allocation of many channels/electrodes to resolution of fundamental frequency (as in §2.6.1.1) leaves fewer channels available for coding of the remainder of the speech/music spectrum. Having fewer channels to code spectral information is likely to impact negatively on the ability to understand speech and identify timbre. Second, F0 place coding potentially introduces interference between the competing demands of coding F0 and the overall spectral shape (or timbre) using place of stimulation. For instance, F0 cues to place-pitch may be affected by variations in higher formant frequencies in speech and vice versa. Third, for some sounds, the amplitude of the fundamental can often be small compared to that of higher harmonics (e.g., for a vowel with a high first formant such as /a/, or a trumpet with high resonance) and thus a relatively weak level of stimulation might result when coding F0 place. While the amplitude of the fundamental could be increased, this may lead to distortions in the coded spectral envelope and hence impact on vowel/timbre perception. Fourth, the limited spectral resolution and selectivity observed in CI hearing (due in the main to the broad spread of the neural activation to electrical stimulation) means that closely spaced partials of F0 may not be resolved despite separation of those partials in the channel stimulus signals. Finally, the relationship between frequency and place in normal hearing is not necessarily preserved in the implanted system where considerable variation in place-pitch order across electrodes both within and between subjects (including some cases where place-pitch reversals are encountered) has been observed. Underlying reasons for this variability include details related to geometry (e.g., position of stimulating electrodes in relation to receptive auditory neurons) and pathology of the impaired cochlea (e.g., number and location of surviving auditory neurons and the condition/impedance of surrounding tissue and fluid in the cochlea). These factors, together with effects of stimulus parameters such as, mode of stimulation, pulse-width, rate and level of stimulation, etc, are likely to impact on the spatial spread of the excitation field (e.g., Cohen, 2009) and consequently on aspects of the perceived place-pitch. For these reasons the present research has instead focused on improving pitch perception by enhancing rate coding of F0 information.
2.6.2 Rate-pitch coding strategies

Some of the earliest CI strategies to code rate-pitch information were the F0/F1/F2 and Multi-peak (MSP) speech-feature extraction strategies (e.g., Seligman et al., 1984, 1992; Skinner et al., 1991). For voiced speech signals, the fundamental frequency was estimated and used to control the stimulation timing at four electrode sites. Two of those sites corresponded in place to estimates of the first two formant frequencies of the vowel and their stimulus intensity was derived from each formant’s estimated amplitude. The other two sites were fixed to two basal electrodes and their stimulus intensity was derived from the envelope output of two band-pass filters spanning the 2-4 kHz range. For unvoiced signals, the four electrodes were activated at a random rate of stimulation and included another more basal fixed electrode site in which stimulus intensity was derived from a 4-6 kHz band-pass filter. While this strategy did provide sufficient information to allow for recognition of most speech sounds in quiet, poorer results were obtained in noisy and reverberant conditions due in part to the difficulty of estimating speech features accurately in noisy conditions. More importantly, compared to more modern clinical strategies which employ a higher constant rate of stimulation to carry amplitude fluctuations in the speech envelope across a larger number of electrodes (e.g., SMSP/SPEAK, CIS, and ACE) the MSP strategy was shown to provide poorer speech recognition performance (e.g., McKay and McDermott, 1993; Skinner et al., 1999). It was shown that the F0-based strategies provided poorer transmission of speech features such as place and manner of articulation. The improved transmission of these speech features by the vocoder-based (SMSP/SPEAK) strategies were attributed to improved coding of spectral and low-frequency (≤ 50 Hz) temporal envelope information. The finer place-pitch coding resolution provided by sequential activation of adjacent electrodes that arises in vocoder-based strategies such as SMSP (McKay et al., 1992; McKay and McDermott, 1996) most likely contributed to the improved coding of spectral information. Furthermore, due to the higher analysis/stimulation rate employed by some of the more recent examples of these strategies (e.g., ACE and CIS), short-duration and rapidly changing speech features in the temporal envelope are better coded. Thus an important consideration in the development of rate-pitch coding strategies is to ensure that sufficient spectral and temporal envelope information is conveyed so that similar levels of speech perception compared to existing clinical strategies is maintained.
2.6.3 Amplitude modulation rate-pitch coding strategies

A number of experimental modifications of modern strategies have been developed that enhance temporal envelope cues to F0 (e.g., Geurts and Wouters 2001; Green et al., 2004, 2005; Vandali et al., 2005; Laneau, 2005; Laneau et al., 2006; Milczynski et al., 2009, 2012). These strategies all coded enhanced F0 temporal information via amplitude modulation of the stimulus envelope. In general, the enhancement of temporal cues provided by those strategies was achieved by: (1) increasing the depth of F0 modulation coded in the stimulus envelope of each channel (§2.3.3.1; refer to McKay et al., 1994, 1995; Geurts and Wouters, 2001 for evidence supporting better rate-pitch discrimination with increasing modulation depth); (2) ensuring that F0 modulation in each channel was mainly presented in phase across channels so that interactions between neighbouring electrodes, due to the spread of the excitation field, did not distort the coded temporal information (§2.3.2.8; see McKay and McDermott, 1996 who examined the integration of temporal information coded on neighbouring electrodes); and (3) for one strategy, employing a modulation function with a rapid onset rather than sinusoidal-like modulation (§2.3.3.3; see Green et al., 2004, 2005, and Landsberger, 2008). It is thought that those steps promoted greater synchrony to F0 in the timing of auditory neural responses to the stimuli.

2.6.3.1 F0 CIS

The strategy examined by Geurts and Wouters (2001) increased the depth of envelope fluctuations in channel signals of the CIS strategy as a means of enhancing temporal cues to pitch, see Fig. 2.8. However, average results for four adult CI recipients in F0 discrimination tests using synthetic vowels with F0s in the range of 149 to 370 Hz, were not significantly different from those of standard CIS. One possible reason for this outcome was that the standard CIS strategy already provided sufficient temporal cues to pitch for the synthetic vowels tested in the study. In addition, because each vowel was synthesised using fixed formant frequencies and bandwidths, place coding would have remained relatively fixed across F0s for each vowel. For natural spoken vowels in which confounding place cues to pitch might arise, test difficulty and reliance on F0 rate cues would be greater (§2.4.1.3; Sucher and McDermott, 2007; Galvin et al., 2008).
2.6.3.2 Sinusoidal and sawtooth modulation of CIS

Green et al. (2004) examined a strategy that enhanced temporal cues to F0 applied to the CIS strategy. Temporal cues were decomposed into two separate components, one consisting of the slow-rate information conveying the dynamic changes in spectral shape, while the second provided F0-related information in the form of a simplified synthesised waveform (sinusoidal, sawtooth, or sharpened sawtooth) that was used to modulate the low-pass filtered (slow-rate) envelope signals in each channel in much the same way as the F0 CIS strategy described above. The F0 information was estimated for each stimulus tested using non-real time techniques and applied to the processed stimuli. The authors noted that for a real-time implementation of the strategy, an F0 extraction process that was sufficiently reliable in conditions of noise was needed. Amplitude modulation was applied to all channels of the CIS strategy by multiplying the slow-rate channel information by the modulation function. The modulation depth applied in each channel was equal to 100% of each subject’s electrical dynamic range (i.e., from maximum comfortable loudness down to the threshold stimulation level). The sawtooth modulation function exhibited a rapid onset at the start of each modulation period followed by a linear decay of current over the modulation period. The sharpened sawtooth decayed to its minimum level by half of the modulation period. These functions presumably provide a better representation of F0 in the neural firing pattern. Group average results in eight adult CI users for identification of glide direction in synthetic diphthongal stimuli were significantly better using the sawtooth modulation functions compared to sinusoidal modulation (which emulated modulation coded by the standard CIS strategy). In addition, results for the sharpened sawtooth condition were equivalent to those in which F0 was coded using pulse rate alone, suggesting that the
sharpened sawtooth function provided close to the maximum temporal cue salience as possible using pulsatile electrical stimulation.

In a follow-up study with four adult CI users, Green et al., (2005) observed similar benefits in question/statement (intonation) identification using the sharpened sawtooth waveform. However, they also found that recognition of vowels and formant discrimination was degraded by this processing and suggested this was due in part to poorer coding of spectral information. They suggested that because the modulation waveform had a relatively narrow peak compared to other modulation functions examined, only a small part of the F0 cycle carried sufficient intensity to convey spectral information. That argument is similar to that presented earlier for the MSP strategy, in which the F0 stimulation rate employed was typically too low to sufficiently convey dynamic changes in the spectral/temporal envelope (§2.6.2). Another possibility is that the deeper/sharper modulation function employed by Green and colleagues may have resulted in greater reductions to channel loudness, particularly for channel intensities nearer to threshold levels of stimulation (e.g., as per the study by McKay and Henshall, (2010), which demonstrated that loudness is inversely proportional to modulation depth). Such reductions in channel loudness may have distorted coding of spectral information. Although, Green and colleagues did not report on any loudness variations between strategies, compensation for such reductions in loudness may be necessary to preserve transmission of spectral and temporal envelope information.

2.6.3.3 F0 modulation enhancement in ACE

Vandali et al. (2005) examined pitch perception using three experimental strategies: Modulation Depth Enhancement (MDE) (Vandali and van Hoesel, 2003; Vandali and van Hoesel, 2004a); F0 Synchronised ACE (F0Sync) (McDermott and McKay, 2005); and Multi-channel Envelope Modulation (MEM) (Vandali et al., 2004b). Those strategies were all modifications of the ACE strategy that enhanced coding of F0 information in the stimulus envelope of each channel (see Fig. 2.9).
Chapter 2: Background and Relevance to Research

Figure 2.9. Stimulus output pattern (electrodogram) recordings for the (a) ACE, (b) MDE, (c) F0Sync, and (d) MEM strategies for a 200 ms portion of the C#3 male /a/ vowel used in the pitch-ranking tests (Vandali et al., 2005). In each electrodogram, time is shown along the abscissa and electrode number along the ordinate. Electrode numbers for the Nucleus 24 implant system begin at 22 for the most-apical electrode, and decrease in number for more-basal electrodes. Each stimulus pulse recorded from the output of the processor is shown as a vertical bar. The heights of the vertical bars represent the stimulus intensity in clinical units for which minimum amplitude corresponds to the threshold, and maximum amplitude corresponds to the maximum comfortable loudness level.

Compared to ACE, those strategies provided deeper modulation cues to F0 coincidentally in time across all activated electrodes. The MDE strategy expanded the depth (peak to trough ratio while holding the peak level fixed) of any modulation in channel envelope signals that were within the frequency range of approximately 80-300 Hz. A second version of this strategy also applied small temporal offsets to channel signals so as to align their temporal envelopes in time. The MEM strategy (Fig. 2.10), smoothed all channel envelope signals and then modulated them by an enhanced (expanded) representation of the temporal envelope of the broadband signal which inherently contained F0 periodicity information. The F0Sync strategy, like that
proposed by Green et al., (2004, 2005) employed a non-real-time F0 estimator to modulate the smoothed channel envelope signals by a pulse train waveform having a duty cycle of approximately 33%. Group average results of pitch ranking tests using sung vowels separated by half an octave, with F0s within the range that could be discriminated using pulse rate, were significantly better than those of ACE for all three experimental strategies.

One of the experimental strategies (MEM) examined by Vandali et al., (2005) was examined further using lexical-tone recognition tests in a group of nine Cantonese-speaking CI users (Ciocca et al., 2005; Wong et al., 2008). Results however failed to demonstrate any significant benefits compared to the ACE or CIS strategies. Possible explanations for that outcome are that: (a) the F0 differences between tonal contrasts, which can be as low as 1-2 semitones (e.g., Bauer and Benedict 1997; Ciocca et al., 2002), were too small to be discriminated by the CI users in the study who obtained average F0-rate DL of approximately 2.5 semitones for a 160 Hz, three-harmonic, complex tone; (b) the rate-pitch mechanism exploited in CIs is too sluggish (§2.3.3.2) to be useful in discrimination of F0 contours in Cantonese tones which can have short durations and can undergo rapid changes in F0 (e.g., Bauer and Benedict 1997); (c) the rate-pitch information coded by MEM, which was derived from the temporal envelope of the broadband signal, was not salient enough and may not have accurately reflected...
Chapter 2: Background and Relevance to Research

F0 due to the broad, and/or variable shape of the modulation waveform, and/or the presence of higher frequency peaks in the temporal envelope for some stimuli; and (d) listeners may have been unaccustomed to attending to F0 rate information for judging tone and confounding cues to pitch coded by place of stimulation may have dominated or disrupted temporal rate-pitch information ($\S 2.3.2.5; \S 2.3.2.7$).

2.6.3.4 F0mod ACE

The experimental strategy evaluated by Laneau et al., (2005, 2006) known as F0mod was also similar to that of Green et al. (2004, 2005). It employed an auto-correlation of the power spectrum to estimate F0. This estimate was used to control the modulation rate applied to the band-pass filtered, slow-rate envelope signals of an ACE-like strategy. A sinusoidal function was used to modulate channels in phase by 100% of each subject’s electrical dynamic range in each channel. For a group of six adult CI users, mean results for F0mod compared to ACE were significantly better for F0 discrimination of synthesised musical notes played by five different instruments for F0s up to 250 Hz. In addition, improved melody recognition of familiar Flemish songs (with rhythm cues removed) was shown. The authors however noted that an assessment of any adverse strategy effects on speech intelligibility was needed and that a more accurate, robust, and efficient F0 estimation technique was required before the strategy could be successfully applied in a clinical system.

More recently, a modification of the F0mod strategy was further assessed in tests of musical pitch perception (Milczynski et al., 2009) and tonal language perception. (Milczynski et al. 2012). The modifications included the addition of a voiced/unvoiced decision criterion to determine when to apply F0 modulation to the band-pass filtered channel signals. It also utilised band-pass filters identical to those used by ACE (the original F0mod strategy employed a higher frequency resolution filterbank with different band-pass filter centre frequencies than those employed by ACE.) For the study reported by Milczynski et al., (2009), improved melodic contour identification and recognition of familiar melodies using synthetically generated stimuli was observed for F0mod compared to ACE in a group of five Nucleus 24 CI users. The authors also observed differences in the perceived loudness of complex harmonic tones processed by each strategy. For low F0s, less than approximately 150 Hz, the stimuli processed by F0mod were noticeably louder than those of ACE, but at higher F0s, the opposite was observed.
Milczynski and colleagues (2012) went on to examine Mandarin word/tone and sentence recognition in quiet and speech-shaped noise in a group of four Mandarin-speaking Nucleus 24 CI recipients. The results showed a modest improvement in lexical tone recognition using F0mod compared to ACE for Mandarin words spoken by a male speaker (at the poorer SNRs). However no differences between strategies were observed for tone recognition in words spoken by a female speaker, in word (and phoneme) recognition scores for male and female speakers, or in sentence perception tests, at all signal-to-noise ratios tested (quiet, +15, +10, +5, 0 dB). While a small benefit using F0mod was observed in tone recognition for low F0s at low SNRs, no improvement in overall word recognition was found. The authors showed that word recognition was more sensitive to perception of vowels rather than tone and suggested that the lack of any difference in word recognition scores could be explained by the similarities in coding of vowel information provided by each strategy, rather than by differences in coding of tonal information. Nevertheless, the authors suggested that with further experience, subjects may learn to better utilise the enhanced F0/tone cues provided by F0mod to perceive speech. It should however be noted that when interpreting outcomes in noise, the results are only indicative of best possible performance using such a strategy because the F0 modulation rate applied to the speech material was derived from the speech processed in quiet. In a real implementation of such a strategy, poorer performance of the F0 estimator would thus be expected in noise which would probably adversely affect tone recognition. In an analysis of the F0 estimator used in F0mod, the authors concluded that it would provide reasonable performance for SNRs down to approximately +10 dB but for lower SNRs an improved F0 estimator would be necessary.

2.6.4 Fine temporal structure coding strategies

Despite limitations in the upper rate that can be coded by CI systems, researchers have examined strategies that code fine temporal detail using pulse rates as high as approximately 1200 PPS/CH. For instance, the Peak Derived Timing (PDT) strategy (van Hoesel and Tyler 2003, van Hoesel et al., 2008) coded fine temporal detail in each channel by producing stimuli corresponding in time and amplitude to positive temporal peaks in the band-pass filtered signals. Results of sung-vowel pitch perception tests however showed little benefit compared to the subject’s normal ACE strategy (Vandali et al., 2005) despite the provision of finer temporal detail. Other examples of strategies
that code finer temporal structure include Musical-L (Fearn, 2001); Half-Wave Gating (Swanson, 2008); Spike-based Temporal Auditory Representation (STAR) (Grayden et al., 2004); Temporal Fine Structure (TFS) and Channel-Specific Sampling Sequences (CSSS) strategies used in MED-EL CI devices (e.g., Schatzer et al., 2010); and the HiResolution strategy used in the Advanced Bionics devices (Koch et al., 2004). None of these strategies have demonstrated significant or substantial improvements in pitch perception for complex harmonic sounds or lexical tones compared to clinical processing strategies. This comes as little surprise given that it is well understood that rate-pitch discrimination deteriorates rapidly for rates beyond approximately 300 PPS and that independent temporal information cannot be coded in nearby channels due to largely overlapping spatial excitation fields. Given this, the focus of the present research is directed towards improved coding and perception of F0 information up to approximately 300 Hz using rate cues provided in the envelope of the stimulus signal.

### 2.6.5 Relevance to Research

For the reasons discussed in sections 2.6.1.3 and 2.6.4, the use of F0 place-coding and fine temporal structure coding strategies to improve perception of pitch were not tested in the present research. The research focused on the development of a strategy in which rate, or rather amplitude modulation rate, was used to encode F0. In addition to those issues summarised in section 2.3.4, the following considerations, based on shortcomings of existing experimental pitch coding strategies (§2.6.3) were addressed.

When coding F0 explicitly via rate or amplitude modulation rate in a sound coding strategy, performance will be highly dependent on the accuracy, robustness to noise, and temporal latency/sluggishness of the F0 estimator. Most of the experimental strategies that have been evaluated thus far have utilised non-real time estimators in which F0 was determined prior to the experiments and so its accuracy could be checked ahead of time and corrected if necessary. Although many varied techniques for estimation of F0 exist (see Hess, 1992; Hermes, 1993; de Cheviné and Kawahara, 2001; for reviews) the computational requirements and complexity of these algorithms impose severe limitations on their implementation in a real-time systems, given the limited processing power of existing portable processors (body worn or BTE). While use of simpler techniques can be achieved in present systems performance is likely to be compromised.

For instance, performance of the auto-correlation based F0 estimator used in F0mod (Milczynski et al., 2009; 2012) degraded at an SNR of around +10 dB and yet many CI
recipients can tolerate and understand speech well for SNRs approaching 0 dB. The need for more robust and practical F0 estimators was thus one of the challenges faced by the present research which is addressed in chapter 4.

When and for which channels F0 temporal information is coded by strategies is also an important design consideration that has not been addressed well by existing strategies. For instance, the most recent implementation of the F0mod strategy does employ a voiced/unvoiced decision criterion to determine whether to modulate the channel signals. An all-or nothing approach is applied to all channels in the system equally. However, because real-world signals often contain a time varying combination of F0 harmonic and inharmonic components distributed across the frequency spectrum, the relative degree of F0 modulation applied should ideally be determined for each channel separately based on the degree to which each channel contains information related to the estimated F0. The experimental strategy developed within the present research employed such an approach (see chapter 4).

While the main objective of F0 rate coding strategies is to improve perception of F0 as a cue to pitch, care must be taken to ensure that the loudness of channel envelope information is coded similarly to that of existing clinical devices. Experiments using amplitude-modulated electrical pulse trains presented at a single electrode site have shown that for a fixed peak level of stimulation, loudness is inversely proportional to modulation depth (§2.3.3.6). Assuming that effect also applies when activating many channels near-simultaneously with the same amplitude-modulated pulse trains, then channel and overall loudness in experimental sound coding strategies is also likely to be affected by the application of deep F0 modulation. For instance, Milczynski et al., (2009) noted that subjects reported differences in loudness between ACE and F0mod when listening to complex harmonic tones. Because sounds continuously vary in their harmonic content, differences in overall loudness will also vary with the amount and depth of F0 modulation applied and so coding of the low-frequency temporal envelope is likely to be distorted. In addition, differences in channel loudness that arise due to selective modulation of some channels are likely to affect coding of the short-time spectral envelope. The present research examined the degree to which channel loudness was affected by parameters of amplitude modulation (see chapter 3) so that appropriate techniques for compensating for differences in overall loudness between clinical and experimental strategies could be developed (see chapter 5).
Chapter 2: Background and Relevance to Research

It is not clear from existing studies with experimental F0 coding strategies as to how well the relationship between coded F0 information and rate-pitch in electric hearing reflects that between F0 and pitch in normal hearing. Studies examining that relationship using unmodulated electrical pulse trains presented at single electrode sites have shown that labelling of pitch intervals with changes in F0 are similar to that reported in normal hearing (§2.3.2.3). However, it has been shown that the pitch-height of amplitude-modulated electrical pulse trains can be higher than that of unmodulated pulse-trains (§2.3.3.4) and so the question of whether and for what conditions the relationship holds for modulated electrical signals remains and is examined in chapter 3.

2.7 Training to improve perception in cochlear implant hearing

There is an extensive body of literature in the field of training that is far too great to be presented here. Instead the following review focuses on background literature and approaches relevant to the specific aims of the present research.

2.7.1 Previous experience

The effect of previous musical experience (training) in normal-hearing listeners on discrimination of frequency was investigated by Spiegel and Watson (1984). Frequency difference limens for musicians were in the range 0.10 to 0.45% in contrast to 0.18 to 1.7% for non-musically trained listeners. Similar outcomes were observed on tone pattern discrimination tests. Micheyl et al., (2006) found that frequency discrimination thresholds for non-musically trained listeners approached that of highly experienced musicians after approximately 4-8 hours of training. Non-musically trained normal-hearing listeners can also find it difficult to discriminate pitch when accompanied by changes in timbre (e.g., Pitt, 1994). Similar effects of timbre have also been seen in studies with CI recipients in which it was observed that musically experienced CI users were generally less confused by timbre when judging pitch than non-musical experienced listeners (e.g., Sucher and McDermott, 2007; Galvin et al., 2008).

2.7.2 Auditory training in normal-hearing listeners

Auditory training has been shown to improve discrimination of many aspects of acoustic signals in music and language. For instance, positive effects of training have been observed for discrimination of lexical tone (e.g., Wang et al., 1999; Wayland and Guion, 2004; Wayland and Li, 2005; Francis et al., 2008) and consonants (e.g., Jamieson and
Chapter 2: Background and Relevance to Research

Morosan, 1986, 1989; McCandliss et al., 2002). Improvements in lexical tone discrimination have also been observed by normal-hearing trainees listening to simulated cochlear implant audio (Smith, 2010). A variety of training paradigms were examined in those studies which included the use of a fixed/closed set of natural sounding stimuli, use of high variability stimuli, use of perceptual fading (exaggerated to natural or diminished stimuli) with adaptive or randomised control of stimulus contrasts, and use of feedback.

2.7.2.1 High variability stimuli

The effect of auditory training on recognition of Mandarin tone by English-speaking adults was investigated by Wang et al., (1999). A high-variability training paradigm (Logan et al., 1991) was employed which involved identification of the four Mandarin tones using a variety of phonetic contexts in natural words produced by a variety of talkers. The training was conducted in eight sessions over a two-week period. Sixteen normal-hearing speakers of American English took part in the study, eight of which participated in the training and eight were controls. Tone recognition tests were conducted prior to training, after the training period, and again six months after the training period. A significant improvement in overall tone recognition for the trainees was observed post training compared to the control group. The improvements also generalised to other test stimuli and to stimuli produced by other talkers. Improved scores for the trainees at six months post training were also retained.

2.7.2.2 Perceptual fading

Jamieson and Morosan (1986) examined the effects of training in ten French-speaking Canadian normal-hearing adults on discrimination of the voiced and voiceless “th” fricative sounds of English. Performance using natural and synthetic speech tokens was measured prior to training and after training. The training was based on an adaptive technique of perceptual fading with feedback which involved identifying synthetic tokens in which acoustic differences between contrasts were exaggerated. As subjects’ performance improved, those differences were reduced towards those of more natural-sounding tokens. After approximately 90 minutes of training, improved discrimination of both synthetic and natural tokens was observed relative to that of an untrained control group. In a follow-up study (Jamieson and Morosan, 1989), a similar procedure was adopted but in this case only a single contrast of the tokens was used during training.
Chapter 2: Background and Relevance to Research

That contrast was chosen to be highly discriminable but not unnatural sounding. While a moderate improvement after training for both synthetic and natural tokens was observed, it was not clear whether the approach was as effective as that of the perceptual fading technique examined in their earlier study.

The effects of training on identification of contrasting /r/ and /l/ stimuli, as in “rock-lock”, by Japanese adults learning English was examined by McCandliss et al., (2002). Two training paradigms were examined each with and without feedback. The first paradigm used an adaptive (perceptual fading) technique in which the contrast between synthetically generated stimuli was initially exaggerated and then adjusted to maintain accurate identification. The second approach used a fixed stimulus contrast that was discriminable by native English speakers but difficult for the Japanese trainees. Tests were conducted prior to training and after a 3-day period of training. Eight subjects participated in each of the four training conditions and eight in the control (untrained) condition. For training with feedback, both the adaptive and fixed approaches provided improvements compared to control subjects and results also translated to benefits for other contrasts (e.g., “road-load”). Without feedback, improvements due to training were only observed for the adaptive training procedure.

Iverson et al., (2005) compared four training approaches in Japanese adults’ perception of the English /r/ and /l/ contrast. Separate groups of subjects received either high variability phonetic training (natural words from different talkers), all enhanced (exaggerated synthetic stimuli), perceptual fading (exaggerated synthetic stimuli which decreased in exaggeration as training progressed), or secondary cue variability (secondary acoustic cues that might disrupt perception of the primary cue held fixed and gradually varied as training progressed). A total of seventy-three subjects completed the testing but only sixty-two were used in the analysis. Positive effects of training were observed for all training approaches but there was no significant difference between techniques. The outcomes suggested that use of high variability stimuli for training can be just as effective as perceptual fading and that listeners’ use of secondary acoustic cues can be important to second language phoneme learning.

The effect of training on perception of Mandarin tone by normal-hearing English speakers was examined by Smith (2010) using various perceptual fading approaches. Only one lexical tone contrast was trained: Tone 2 (rising) versus Tone 3 (dipping and rising). Tone contrasts were presented within monosyllabic words (either one or four) spoken by a female speaker (F0 range ~160 to 260 Hz). The contrast between tones was
exaggerated or diminished by adjusting their mean F0 and tone duration to produce a set of 28 perceptually faded tones. Two training schedules were examined. The first employed an adaptive procedure to systematically adjust the contrast between stimuli based on performance. The second randomly selected stimuli to compare from those in the set of 28 perceptually faded stimuli. In each trial a pair of stimuli was presented acoustically and the listener indicated whether the pair were the same or different. Training consisted of 128 trials and no feedback was provided after each trial. Assessment tests using similar stimuli were conducted pre- and post-training. Ninety-six subjects participated in the study. Half of them completed the experiments listening to a CI-simulation of the stimuli and the other half listened to the normal stimuli. For each group of subjects, half used the adaptive training schedule and the other half the random schedule. While some improvement in performance was seen after training, no significant effects of training were found for subjects trained with the normal or the CI simulated tones, or for each training schedule, suggesting in general that the perceptual training was not effective without feedback.

The experiment was repeated with the addition of visual feedback regarding the correctness of each response. A smaller group of subjects (forty-five) participated and only the CI-simulated tones were used. A significant improvement in pre- to post-training performance was observed for both training schedules. A second post-training test was conducted five minutes after the initial post-training test to examine the robustness of performance over time. Better retention was observed for the randomised training schedule compared to the adaptive schedule, however, because different measures of performance were used for each schedule, the results could not be directly compared. Overall, based on the relative performance of each training schedule across tests, the author concluded that the adaptive approach led to greater improvements over the course of the training compared to the random schedule, but after a short break, the randomised approach was the only schedule to show a definite and persistent improvement in perception. The implications of those outcomes to training approaches are however unclear due to the absence of data from untrained control subjects and the relative short period of training used.
Chapter 2: Background and Relevance to Research

2.7.3 Auditory training in cochlear implant recipients

Positive effects of training on the ability to identify instruments (e.g., Gfeller, et al., 2002c) and melody (e.g., Gfeller, et al., 2000b; Galvin et al., 2007) have been observed in adult CI recipients when using clinical coding strategies.

2.7.3.1 Melodic contour identification training

Galvin et al., (2007) examined the effects of training on perception of melodic contours using the Melodic Contour Identification (MCI) test. This test (and the training paradigm) consisted of a closed set melodic contour identification test with feedback provided in the training component. The complex harmonic tones used in the MCI test included sine-phase harmonics of the fundamental frequency and its first two harmonics. The relative level of F0, 2×F0, and 3×F0 was 0, -3, and -6 dB respectively. The stimuli consisted of nine melodic contours (rising, rising-flat, rising-falling, flat-rising, flat, flat-falling, falling-rising, falling-flat, and falling) composed of five tones (see Fig. 2.11), separated by F0 intervals ranging from 1 to 5 semitones.

![Figure 2.11. Melodic contours used in MCI test](image)

For MCI testing, the root (base) note in all contours was A3 (220 Hz), A4 (440 Hz), or A5 (880 Hz). A total of 135 melodic contours were available in the MCI test (9 melodic patterns × 5 semitone distances × 3 root notes). For MCI training, one root note within the range 220-880 Hz, excluding A3, A4, or A5, was randomly selected, and one F0 interval between notes was selected per training run. In each training run, 25 melodic
contours (distributed randomly within the stimulus set) were presented. One contour was randomly selected (without replacement) from the set and played to the subject. The subject then selected one of the possible responses shown on a screen. The number of possible responses varied from 2 to 9 depending on the level of difficulty. If the subjects answered correctly, visual feedback was provided and a new contour was presented. If however they answered incorrectly, visual and audio feedback was provided in which the correct response and the subject’s incorrect response were repeatedly played back for comparison. When scores of 80% or higher were achieved within a training run, either the level of difficulty was increased and/or the F0 interval between notes was reduced.

Six adult CI subjects participated in the training study. MCI and Familiar Melody Identification (FMI) tests were conducted at the start of the study. The subjects then underwent training in which five of the subjects trained for half-hour per day, every day for a period ranging from 1 week up to nearly 2 months depending on the subject. For the sixth subject, 3 hours per day of training was provided over a 5-day period. MCI tests were conducted at regular intervals during the training period and at its completion. The FMI test was repeated at the completion of the training period in four of the six subjects. Results showed that MCI scores improved significantly from 50% before to 78% after training. In addition, the effects of training translated to improvements in FMI scores both with and without rhythmic cues, although benefits of training were only significant in FMI scores without rhythmic cues (mean pre-training scores were 52.4 and 16.7% with and without rhythmic cues respectively which increased to 61.5 and 37.4% respectively). However, while these data demonstrate that benefits to pitch and melody perception can be obtained via training, the absence of comparative control data for subjects that did not undergo training weakens the validity of the outcomes. In addition, given that the stimuli used in the training and testing consisted only of synthetic complex harmonic tones in which rate and place cues to pitch were presumably in agreement with changes in F0, the possibility exists that similar benefits of training might not be obtained when listening to more realistic/natural musical sounds due to adverse effects that variations in spectral timbre (place coding) can have on judgement of pitch (§2.3.2.5; §2.3.2.7; §2.4.1.3; §2.4.3.3).
2.7.3.2 Melody recognition training

A computerised auditory musical training program was developed by Gfeller et al., (2000b) which consisted of 48 modules (each approximately 30 minutes in length) of self-administered training in which subjects were expected to complete four modules per week over a 12-week period. The training material comprised twelve simple melodies (such as “Happy Birthday”) and eighteen complex songs from three musical genres (classical, country and western, and pop). The simple melodies were presented as melody-only or as melody-plus-harmony without lyrics and were repeated ten times over the 12-week period. The melody was accompanied by its title and a visual representation of changes in pitch and rhythm on the computer screen. Each week, the final lesson included a brief review of the material presented and subjects were tested on their discrimination and recognition of the melodies from a closed set of options. The complex songs consisted of 3-minute excerpts presented 11 times over the 12-week period. Each excerpt was accompanied by a brief written lesson in which subjects were required to focus attention during listening on answering questions related to aspects of the music.

The effects of the training program on melody and song recognition was examined in a group of Nucleus adult CI recipients using their clinical processors (Gfeller et al., 2000b). Eleven subjects undertook the training program and nine acted as controls that experienced incidental exposure to music in their normal daily routine during the 12-week period. For both groups of subjects, recognition and appraisal of familiar melodies (as per the tests described by Gfeller et al., 2002a referred to in §2.4.2) and complex songs excerpts was conducted prior to the 12-week period and at the end of that period. For the familiar melodies test, a significant improvement in recognition was observed for the subjects who participated in the training program where mean scores increased from 12.5% pre-training to 23.4% post-training. In contrast, no significant improvement in scores was observed for the control subject group (5.3% pre-training and 7.7% post-training). An even greater benefit for the trainee subjects was observed in the complex song excerpt recognition test in which scores increased from 3.7% to 36.5% compared to 2.6% to 6.2% for the control group. Appraisal of the complex songs was also examined using measures of “liking” and “complexity” rated on a scale of 0 to 100 where higher numbers indicated greater liking and higher perceived complexity respectively. For the trainee group, a small but significant improvement post-training was observed with an
increase in the liking rating from 56.1 to 62.5 and a decrease in complexity rating of 42.8 to 38.2. For the control group, no significant effect of training was observed with pre- and post-training liking ratings of 53.0 and 51.2 respectively and complexity ratings of 40.6 and 41.0 respectively.

**2.7.3.3 Instrument identification training**

The effect of the training paradigm described above (§2.7.3.2) on subjects’ ability to discriminate musical instruments was examined in the same group of Nucleus adult CI recipients (Gfeller et al., 2002c). The training consisted of 48 lessons (each around 10 minutes in duration). Four lessons were conducted each week over a 12-week period. The training involved listening examples and exercises for timbre recognition using a variety of instruments, which included those from the string, woodwind, brass, and pitched percussion families (namely the piano). Assessments were conducted pre- and post-the 12-week training period using instrument timbre recognition and appraisal tests. The stimuli used in these tests consisted of a fixed melodic pattern produced by eight familiar instruments which spanned a range of F0s from 131 to 1068 Hz. In the recognition test, subjects were required to identify the instrument from a set of 16 possible instruments. For the timbre recognition test, a significant improvement of approximately 20 percentage points for the trainee group was observed post-training whereas scores for the control group were not significantly different across the 12-week period. Overall pleasantness was also rated significantly higher after training by the trainees but not by the control subjects.

**2.7.4 Relevance to research**

A major aim of training is to promote long-term improvements in performance which also translate to improvements in related tasks or altered contexts (Schmidt and Bjork, 1992). Musically trained NH and CI listeners show better ability to discriminate pitch and to do so independently of changes in timbre as compared with untrained listeners (§2.7.1). This suggests that experience gained through training programs can lead to improved performance on those measures.

Features common to successful training paradigms indentified by Smith (2010) include: (1) the use of high variability stimuli (§2.7.2.1), e.g., use of multiple talkers and varied phonetic content rather than a single talker and fixed phonemes; (2) perceptual fading (§2.7.2.2) beginning with the most extreme exemplars of stimuli and progressing
to more natural sounding exemplars as perception improves, rather than training only with fixed, or natural sounding, exemplars; and (3) the provision of feedback (§2.7.2.2). In addition, for training based on perceptual fading, adaptive (easy to hard) control of exaggeration can lead to improved results during training but randomised control of exaggeration can provided a more robust outcome after training.

For CI recipients, improved perception of pitch, melody, or instrument type has been obtained using training paradigms in which a fixed (closed) set of stimuli have been employed with feedback (§2.7.3.2; §2.7.3.3). High variability (closed-set) training programs have also been examined which demonstrated performance improvements. Those programs employed slow adaptation (e.g., across sessions or blocks) to harder sets of stimuli as subjects perception improved (§2.7.3.1) rather than a strict perceptual fading approach in which adaptation occurs from one trial to the next. In the present research, various elements of the above training techniques were adopted in the development of a pitch training program and schedule which is described in chapter 6.
Chapter 3: Pitch and Loudness of Amplitude Modulated Stimuli

3 Pitch and Loudness of Amplitude-Modulated Stimuli

The pitch elicited by unmodulated and amplitude-modulated electrical pulse trains was examined with six adult cochlear implantees. In addition, for three of those subjects who had some residual acoustic hearing in their contralateral ear, the pitch of unmodulated electrical pulse trains was compared to that of complex harmonic acoustic tones. In the first experiment, pulse rate discrimination and the effects of place and level differences on pitch were examined for unmodulated pulse trains. General results were consistent with previous studies showing that variations in pulse rate, while holding loudness fixed, elicit changes in pitch at low rates, but become progressively harder to discriminate as rates approach approximately 300 pulses-per-second. Variations in place or level of stimulation generally produced changes in pitch consistent with tonotopic place and spread of excitation. In the second experiment, pitch and loudness of unmodulated pulse trains were compared with those of amplitude-modulated stimuli as a function of modulation depth, rate, and shape, and presentation level. The pitch elicited by an amplitude-modulated pulse train was generally higher than that of an unmodulated pulse train with a pulse rate equal to the modulation rate, and generally decreased towards that of the unmodulated pulse train as modulation depth or rate increased, or as presentation level decreased. Sharper/narrower modulation produced lower pitch. In the final experiment, the pitch heights of acoustic complex harmonic tones and unmodulated pulse trains were compared. When electrical pulse rate was equal to the fundamental frequency of the acoustic tone, similar pitch heights were elicited. The results from these experiments indicate that F0 rate-pitch derived from the temporal envelope in existing clinical cochlear implant strategies may often be higher than that of acoustic harmonic tones at the same F0 in normal hearing, and that pitch growth with increasing F0 may be shallower. The relationship between F0 and rate-pitch is expected to be more similar to acoustic stimulation for low F0 rates when using new pitch coding strategies that code F0 information via deep (narrow) amplitude modulation of the stimulus envelope. Although that similarity reduces as F0 approaches the upper limit of rate-pitch discrimination, that limit is reached sooner for the shallow (or broad) modulators used in existing clinical strategies.
Chapter 3: Pitch and Loudness of Amplitude Modulated Stimuli

3.1 Introduction

The literature review in chapter 2 covered much of the background for the present study but a brief summary of the most relevant sections will be revisited here for the purposes of continuity and clarity.

While a high degree of speech recognition in quiet can be obtained by users of cochlear implant (CI) systems, perception of musical pitch and tone in language are generally far from satisfactory (§2.4.1; §2.5). The poor spectral resolution (§2.3.1) and upper limit of rate-pitch perception (§2.3.2) in CI systems restricts coding of fine spatio-temporal information in resolved harmonics of the fundamental frequency (F0) from which pitch is derived in normal hearing (§2.2.1). F0 information is predominately coded by way of amplitude modulation of the stimulus envelope of CI systems. The depth and shape of this modulation is highly dependent on acoustic properties of the signal and characteristics of the sound coding strategy and thus the salience (i.e., pitch strength) and accuracy (i.e., pitch height) of rate-pitch information can vary substantially across signals, acoustic environments, and coding techniques. Considerable research has focused on development of coding techniques to improve rate-pitch salience (§2.6.3) but little has been done to examine its accuracy. The main aim of the present study was to examine the effects of parametric variations of amplitude-modulated electrical pulse trains on pitch height through comparison with the pitch of unmodulated pulse trains. The reason for using unmodulated pulse trains as reference is that it has been shown that accurate identification of musical pitch intervals can be obtained by CI recipients on the basis of changes in pulse rate, at least for low rates (§2.3.2.3).

Pitch percepts arising from temporal stimulus properties presumably reflect differences in neural response properties. For low-rate electrical pulse trains of a few hundred pulses-per-second, a high degree of phase locking and entrainment of neurons to each pulse is observed (§2.3.2.1). However at higher pulse rates, entrainment decreases and population responses exhibit poorer phase locking than that seen at lower rates. That behaviour arises because immediately after firing, neurons are prevented from discharging again during an absolute refractory period, which is followed by a period of relative refraction in which their firing probability is gradually restored (e.g., Shepherd and Javel, 1997; Miller et al., 2001; Shepherd et al., 2004). Hence with increasing pulse rate at a fixed pulse level, discharge rates of neurons progressively saturate in accordance with their refractory behaviour. In addition, because individual
neurons saturate at different rates, the variance in response rates increases (e.g., Miller et al., 2001).

In behavioural studies, it has been well established that changes in pulse rate of an electrical pulse train can elicit different percepts of pitch for rates up to approximately 300 pulses-per-second (PPS) (§2.3.2). While some researchers have reported that rate increases up to as high as 1000 PPS can be detected by some subjects, those results are not necessarily based on discrimination of changes in pitch per se. The progressive increase in pulse rate discrimination thresholds with increasing rate is in good agreement with refractory limitations and increased variance seen in the physiological data. The correspondence between behavioural and physiological outcomes can be modelled by assuming that pitch is derived from a weighted average of the neural response times (e.g., van Wieringen et al., 2003).

Previous studies have examined the relationship between changes in electrical pulse rate and pitch intervals (Pijl and Schwarz, 1995; McDermott and McKay, 1997; §2.3.2.3). Pijl and Schwarz (1995) asked CI subjects to indicate whether the pitch interval between two sequentially presented stimuli that differed only in pulse rate were consistent with their memory of a specific musical intervals. For the three subjects tested, their judgments of musical intervals on the basis of pulse rate were consistent with the same ratios that normal-hearing listeners would associate with changes in F0. Similar results were observed by McDermott and McKay (1997) who examined both musical interval perception and production in a single CI subject that had trained as a musical instrument tuner prior to deafness. Those results suggest that the ratiometric relation between pulse rate and pitch in electric hearing is similar to that between F0 and pitch in normal hearing. However, they do not describe the absolute relationship between the pitch produced by electric pulse rate and acoustic F0. Blamey et al. (1996) compared absolute pitch heights of acoustically presented pure tones and unmodulated electrical pulse trains in adult CI users who had some residual hearing in their non-implanted ear (§2.3.2.4). Although, the authors reported that for matched pitch the average electric pulse rate was approximately equal to the pure-tone frequency, there was large variability in results between subjects, and in general, pitch increased more rapidly as a function of pulse rate than as a function of acoustic frequency. It was suggested that the latter could be explained by a decrease in rate-pitch salience with increasing pulse rate leading to an increase in the contribution of place of stimulation to pitch. Place of stimulation can convey a sensation of pitch according to a tonotopic
arrangement such that stimulation of neurons adjacent to basal electrodes generally elicits a higher pitch than stimulation of apical electrodes (§2.3.1).

Effects of place of stimulation on rate-pitch have also been observed by other researchers (e.g., Pijl, 1997a; Zeng, 2002; §2.3.2.5). Pijl (1997a) used a pitch matching test to examine the effects of place of stimulation on rate-pitch of unmodulated pulse trains in three CI subjects. In general, at low pulse rates (≤ 100 PPS) for which the salience of rate cues is expected to be strong, results demonstrated little effect of place on pitch although some variability was observed. At higher pulse rates, for which the salience of rate-pitch cues is expected to be weaker, the effects of place of stimulation were more significant and in general basal shifts in the place of stimulation produced higher pitch percepts. Zeng (2002) employed a pitch magnitude estimation test to examine the effects of place and rate of unmodulated pulse trains on pitch in four CI recipients. In general for low pulse rates, pitch was dominated by rate and little effect of place was observed. Pitch increased monotonically with increasing pulse rate up to approximately 300 PPS and saturated at higher rates. At those higher rates, pitch was dominated by place and more basal stimulation elicited higher pitch.

Changes in stimulation level have also been shown to influence pitch although there are considerable differences between outcomes of studies (§2.3.2.6). For the small number of subjects examined by Pijl (1997a), general outcomes demonstrated a decrease in pitch with increasing stimulation level, particularly when the rate-pitch percept was weak (i.e., at higher pulse rates). That result was attributed to an apical shift of the low-frequency edge of the spatial excitation pattern with increasing stimulation level. In a group of sixteen CI users examined by Arnoldner et al. (2008), a significant effect of level on pitch was observed in approximately 87% of subjects, with 73% of them indicating a lower pitch with increasing level and the other 27% reporting the opposite. More recently, Carlyon et al. (2010a) observed that small changes in level can have substantial effects on pitch which are not entirely consistent with the effects of stimulation level on spatial excitation patterns. Using a pitch ranking test, in sixteen out of twenty-one cases in which an effect of level was observed, pitch increased with level, which was attributed to increased neural discharge rates at higher levels.

Although earlier CI sound coding strategies, such as Multipeak (§2.6.2), did employ stimulation rates corresponding to F0, their lower overall pulse rate was shown to be partly responsible for reduced transmission of spectral and temporal envelope information in speech compared to subsequent filterbank-based strategies using higher
rates (e.g., McKay and McDermott, 1993; Skinner et al., 1999). Assuming that recognition of speech is most important to CI users, strategies that employ moderate-to-high stimulation rate are therefore preferable. Those higher rate strategies code F0 temporal information through amplitude modulation of the stimulus envelope. Like pulse rate, changes in amplitude modulation rate of electrical pulse trains have also been shown to elicit different pitch percepts (§2.3.3). However, the salience of those percepts can be weaker than those elicited by changes in pulse rate particularly for shallow modulation depths. Furthermore, coding of F0 modulation by current clinical strategies has been shown to be poor mainly due to inconsistent presentation of deep modulation across channels and signals (e.g., Geurts and Wouters, 2001; McDermott, 2004; Vandali et al., 2005). Reduced salience and poor coding of F0 modulation may in part account for the unsatisfactory results of pitch and tonal language perception using clinical devices (§2.4.1; §2.5), although the lack of fine-spatio temporal coding in CIs is also a major contributor.

A number of experimental strategies have been developed that enhance coding of temporal envelope cues to F0 (§2.6.3). Those strategies have been shown to provide significantly better pitch discrimination when compared to clinical strategies for a limited range of F0s (up to approximately 300 Hz). The improvement is thought to be related to provision of deep F0 modulation presented in-phase across stimulation channels. Supporting evidence can be drawn from studies demonstrating improved modulation rate-pitch discrimination with increasing modulation depth (§2.3.3.1). In addition, the benefit of presenting modulation in-phase across channels is in agreement with studies showing that temporal information is integrated across nearby electrodes (McKay and McDermott, 1996) due to the spatial spread of electrical excitation at each site (e.g., Cohen, 2009). However, while those experimental strategies can improve listeners’ abilities to hear changes in F0, they may not necessarily improve the correspondence between F0 and pitch that is experienced in normal hearing.

While under conditions of fixed stimulation place and level, the relative relationship between pulse rate and pitch in electric hearing can be similar to that between F0 and pitch in acoustic hearing, the conditions for which that relationship holds for amplitude-modulated pulse trains (such as those used to encode F0 in CI strategies) are less clear. Previous studies have shown that the pitch elicited by modulation rate is generally higher than that derived from unmodulated pulse rate (§2.3.3.4). McKay et al., (1995) found that for shallow modulation depths, the pitch of a
modulated stimulus was substantially higher than that of an unmodulated pulse train with a pulse rate equal to the modulation rate, but decreased towards that of the unmodulated pulse train as the depth of modulation was increased. For some of the subjects and stimuli examined, the modulation depth required to match the pitch height of the stimuli to within 1 semitone was typically larger than the subject’s electrical dynamic range. It is thus likely that for strategies in which F0 is coded by amplitude modulation, a distorted and inconsistent relationship between F0 and pitch may result from coding of shallow and variable modulation depth. Some experimental strategies have employed a sharpened (i.e., sharp onset and a short duty cycle) modulation function to encode F0 (§2.3.3.3; §2.6.3.2). Although the ability to discriminate pitch was improved using those modulation functions, effects on pitch height were not assessed. Similarly, data describing the effects of modulation depth and rate on pitch height (i.e., McKay et al., 1995) are only available for relatively low stimulation rates. Further research is therefore required to determine the effects of modulation depth, shape, and rate on pitch height.

The present research expands on previous studies by systematically examining the effects of modulation depth, shape, and rate, as well as presentation level, on pitch height and loudness in experiment 2. Stimulus parameters examined were chosen to be relevant to recent coding strategies that employ moderate to high stimulation rates. Pitch was always compared to that of an unmodulated pulse train for which it has been shown that accurate identification of musical intervals can be obtained on the basis of electrical pulse rate. In experiment 1, tests were conducted with unmodulated pulse trains to establish base-line performance of each subject on measures of rate-pitch discrimination and sensitivity to changes in place and level of stimulation. Those data aided in interpretation of the modulation rate-pitch results from experiment 2, which included variations in stimulation level and hence in the spatial spread of excitation. In experiment 3, the pitch heights of unmodulated electrical pulse trains were compared to those of acoustic tones for CI subjects who used a contralateral hearing aid. The main aim of that experiment was to examine the absolute relationship between electric rate-pitch and acoustic pitch, albeit with a small subject pool.

The outcomes of the research are directly applicable to the development and optimisation of experimental rate-pitch coding strategies such as the Enhanced-Envelope-Encoded Tone (eTone) strategy developed in the present research (see chapter 4). For strategies that code F0 information using amplitude modulation, the relevant
Chapter 3: Pitch and Loudness of Amplitude Modulated Stimuli

Questions examined in this study include: (1) how deep does the modulation need to be to provide a similar pitch to that elicited by unmodulated pulse rate at F0; (2) at that modulation depth, is coding of loudness affected to the extent that it would need to be corrected; (3) does the effect of modulation depth on pitch vary with modulation frequency or presentation level; and (4) can a sharpened modulation function elicit a more similar pitch height to that of an unmodulated pulse train than a sinusoidal modulation function?

3.2 General Methods

3.2.1 Subjects

Six adult users of the Nucleus 24 cochlear implant system took part in this study. Their demographic details are listed in Table 3.1. All had at least 1.5 years of experience with their CI device. Subjects S1, S2, S4, and S5 had no formal musical experience. S3 and S6 practiced music (piano and guitar respectively) prior to the onset of deafness. S6 continued to practice music post implantation, whereas S3 did not. S5 is a speaker of tonal languages and is thus experienced in tone-discrimination. All subjects indicated that they had enjoyed listening to music prior to deafness.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Gender</th>
<th>Age (yr)</th>
<th>Years implanted</th>
<th>Musical/Tonal experience</th>
<th>Cochlear Implant</th>
<th>Clinical Processor</th>
<th>Clinical Strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>M</td>
<td>78</td>
<td>1.5</td>
<td>None</td>
<td>CI24RE (CA)</td>
<td>CP810</td>
<td>ACE</td>
</tr>
<tr>
<td>S2</td>
<td>M</td>
<td>65</td>
<td>12</td>
<td>None</td>
<td>CI24M</td>
<td>Freedom</td>
<td>SPEAK</td>
</tr>
<tr>
<td>S3</td>
<td>F</td>
<td>66</td>
<td>2.5</td>
<td>Music prior to CI</td>
<td>CI24RE (CA)</td>
<td>Freedom</td>
<td>ACE</td>
</tr>
<tr>
<td>S4</td>
<td>M</td>
<td>85</td>
<td>3</td>
<td>None</td>
<td>CI24RE (CA)</td>
<td>CP810</td>
<td>ACE</td>
</tr>
<tr>
<td>S5</td>
<td>M</td>
<td>55</td>
<td>12</td>
<td>Tonal language speaker</td>
<td>CI24M</td>
<td>ESPrit3G</td>
<td>ACE</td>
</tr>
<tr>
<td>S6</td>
<td>M</td>
<td>62</td>
<td>3</td>
<td>Music prior and post CI</td>
<td>CI24RE (CA)</td>
<td>CP810</td>
<td>ACE</td>
</tr>
</tbody>
</table>

Table 3.1. Subject details: gender, age, number of years implanted at commencement of the study, musical/tonal-language experience prior to deafness onset, implanted cochlear implant device (CA = Contour Advance; otherwise straight array), clinical speech processor and coding strategy.

3.2.2 Electrical stimuli

For all electrical stimuli, electrodes were stimulated in monopolar configuration between an intra-cochlea and extra-cochlea reference electrode (MP1+2 for the Nucleus 24 system). Biphasic, cathodic leading, current pulses of 25 μs per phase with an 8 μs inter-phase gap were employed. The duration of each stimulus was 300 ms and sequentially
Chapter 3: Pitch and Loudness of Amplitude Modulated Stimuli

Presented stimuli were separated by 300 ms. For amplitude-modulated pulse trains, modulation rates of 100, 200 and 300 Hz and a stimulation rate of 1800 PPS were used. The stimulation rate was an integer multiple of all three modulation rates and was sufficiently high to ensure that the modulation function was well sampled (§2.3.3.5). The highest pulse level in the modulated pulse trains was fixed at the desired stimulation level and amplitude modulation was applied such that the lowest pulse level was lower than that by 12.5, 25, 50 or 100% of the subjects’ electrical dynamic range (EDR) in clinical units (CUs). In the Nucleus 24 system, CUs range from 0 to 255 which correspond to current levels of 10 to 1750 µA where each step produces an increase in current of approximately 2% or 0.176 dB. Unless otherwise specified, stimuli were presented at a single site (electrode 18) which corresponded to the band-pass filter channel with a centre frequency of 750 Hz in all subjects’ clinical processors. Stimuli were generated on a Windows-based PC using custom software that interfaced to a research sound processor (SPEAR3).

3.2.3 Acoustic stimuli
Complex harmonic tones were constructed from low-pass filtered impulse trains. The impulse trains were oversampled at $8 \times 44100$ Hz and down-sampled to 44,100 Hz. A 10th order low-pass Butterworth filter with a cut-off frequency of 1200 Hz was applied to band-limit the stimuli. The duration of stimuli was 300 ms and their onset and offset envelopes were smoothed by a 20 ms raised-cosine function. All stimuli were presented in a sound-attenuated room via a loudspeaker placed 2.0 m from the subject at an azimuth of zero degrees.

3.2.4 Threshold and comfortable loudness levels
Threshold (T) and comfortable loudness levels (C) of stimulation were measured for each subject for pulse rates of 100, 200, 300 and 1800 PPS on electrode 18. Threshold levels were measured using a one-up one-down staircase procedure (initial step size of 5 CU for the first turning point thereafter reduced to 2 CU; threshold determined from average of last four of six turning points). Comfortable levels were determined by increasing the stimulation level until subjects reported that loudness was similar to that produced by stimulation on a single electrode at the C-level used in their clinical processor. Those levels were then compared between the lowest rate and each higher rate (using an A-B loudness comparison procedure with each presentation order repeated
twice) and adjusted if necessary to achieve optimal loudness balance. The measured T-Levels and C-Levels are presented in Fig. 3.1 in CU. For most subjects T- and C-Levels decreased with increasing pulse rate but by a greater amount for T-Levels. That results in a larger EDR at the higher pulse rates, which is consistent with trends observed in the general CI population (e.g., Kreft et al., 2004). An unusually small EDR at pulse rates of 100 to 300 PPS but not at 1800 PPS was observed for S2. For subject S4, measurements of C-Levels were compromised by facial nerve stimulation with increases in level and rate, so that those C-Levels may not have been accurately balanced across rates shown in Fig. 3.1. However for that subject, those slight differences in loudness across rates did not impact on relative measurements of pitch and loudness performed for each modulation rate examined in experiments 2 and 3.

Figure 3.1. Threshold (T) and comfortable loudness levels (C) of stimulation in clinical units for 300 ms long pulse trains presented on electrode 18 at pulse rates of 100, 200, 300 and 1800 PPS for each subject and averaged across subjects. C-Levels are plotted as upward pointing triangles and T-Levels as downward pointing triangles. The error bars indicate standard deviation of the subject mean.

### 3.2.5 Loudness balancing

An adaptive two-interval, two-alternative, forced choice (2I-2AFC) procedure similar to that described by Levitt (1970) was employed to loudness balance stimuli. The stimuli consisted of a reference and a probe that differed initially in loudness (according to subjective comments) and were presented sequentially in randomised order. The subject was asked to nominate which of the two was loudest. After each response, the level of the probe was adjusted using a one-up one-down staircase procedure that continued until twelve turning points of the probe level had been obtained. An average of the last six turning points was used to determine the matched level for a run. The initial step size was typically set to approximately 15-20% of the subjects’ EDR in CU for the probe.
Chapter 3: Pitch and Loudness of Amplitude Modulated Stimuli

stimuli and was halved after the first three turning points. The final loudness-matched level was determined from an average of at least two consecutive runs of the procedure. Between consecutive runs of the test, the initial level of the probe alternated from being quieter and then louder than the reference.

3.2.6 Data analysis

For each experiment, unless otherwise specified, the data were subjected to analysis of variance (ANOVA) using a general linear model (GLM) in which subject was treated as a random (block) factor and all interactions between main factors were examined. Post-hoc Bonferroni tests \((p < 0.05)\) were used to determine specific treatment differences. All reported means were adjusted for any missing data points in the GLM.

3.3 Exp. 1: Pitch of unmodulated electrical stimuli

The ability of subjects to rank pitch based on pulse rate was measured for unmodulated pulse trains at reference rates of 100, 200, and 300 PPS. In addition, the influence of electrical place and stimulation level on the judgment of rate-pitch was examined using pitch ranking tests in which electrode place was varied by up to \(\pm 1\) or \(\pm 4\) sites or stimulation level was varied by an average of up to -19 to +32\% of the subjects’ electrical dynamic range. Those tests served to determine the extent to which individual subjects could utilise pulse rate as an independent/exclusive cue to pitch and for relating that performance to results of experiments 2 and 3.

3.3.1 Exp. 1a: Pulse rate difference limens

3.3.1.1 Methods

Rate difference limens (DLs) for unmodulated electrical pulse trains presented at a single site (electrode 18) were measured at three reference rates of 100, 200, and 300 PPS. The test employed a two-interval, two-alternative, forced choice (2I-2AFC) procedure in which subjects were asked to nominate which of two randomly ordered stimuli differing in pulse rate was higher in pitch. No feedback was provided. Prior to the test, a set of five stimulus pairs was chosen\(^1\) from a set of eight that differed in rate by \(\pm 0.125, 0.25, 0.5, 1, 2, 3, 4,\) and 5 semitones (or approximately \(\pm 0.7, 1.4, 2.9, 5.8,\)

\(^1\) An informal test was used to determine the smallest rate-interval for which the subject could consistently rank rate-pitch correctly. That rate-interval and four consecutively smaller intervals were chosen for use in the rate DL test.
Chapter 3: Pitch and Loudness of Amplitude Modulated Stimuli

11.6, 17.4, 23.3 and 29.3%) around the reference rate. Note, that the term semitone is used here and throughout the text to describe a frequency ratio that corresponds to the twelfth root of two. It is not intended to imply that a change in electrical rate by that ratio that will produce a change in pitch corresponding to a semitone as defined in the Western musical scale. Within each test run, the five stimulus pairs were presented ten times in randomised order of stimulus pair and pulse rate. For each reference rate, at least two test runs were administered. The level of each stimulus was loudness balanced against a pulse train presented at the measured C-level for each reference rate using the procedure described in Sec. 3.2.5. To discourage the systematic use of loudness when judging pitch, the intensity of each stimulus presentation was roved in steps of 1 CU by up to approximately -10% of the EDR (in CU) or -2 CU, which ever was greatest. Response bias-corrected percent-correct scores (Macmillan and Creelman, 1991) were calculated for each pulse rate interval (stimulus pair) and used to fit cumulative Gaussian functions using weighted linear regression. Thresholds were determined for a performance criterion of \( P_{\text{max}} = 76\% \) corresponding to \( d' = 1 \). Standard deviations of the thresholds were estimated using a bootstrap procedure (Foster and Bischof, 1991).

3.3.1.2 Results and discussion

Pulse rate DLs (expressed in units of semitones) as a function of rate are plotted in Fig. 3.2 for each subject and for the subject group. Average rate DLs (adjusted for the missing data point for S1 at 300 PPS) of 2.4, 3.4, and 4.9 semitones were observed for rates of 100, 200, and 300 PPS respectively. Expressed as a percentage of the pulse rate, for which e.g., 5.9% corresponds to 1 semitone, the average DL for 100 PPS was 14.9% with a range of 4.7 to 53.3%. Subjects S4, S5 and S6 obtained the lowest DLs which were around 1 semitone at 100 PPS and increased to around 2 semitones at 200-300 PPS. DLs for subjects S2 and S3 were around 2-3 semitones at 100 PPS and increased to around 5-8 semitones at 300 PPS. Subject S1 performed worst and had trouble ranking the pitch of the stimuli consistently across runs. His DLs were around 7-9 semitones at 100-200 PPS and were beyond the range of rate intervals examined in this test at 300 PPS.

---

2 For three of the six subjects, response bias was evident when the rate interval was too small to be ranked correctly.
Chapter 3: Pitch and Loudness of Amplitude Modulated Stimuli

Overall, these subject-averaged thresholds are somewhat higher than observed in previous studies in which an average rate-pitch DL of 7.3% with a range of around 2 to 18% at a rate of 100 PPS was reported in a review by Moore and Carlyon (2005). However excluding S1, who had difficulty performing the test, from the analysis, yields a more comparable average DL of 8.8% and a range 4.7 to 17.6% at 100 PPS. A two-way ANOVA (with subject as a random factor) revealed a significant effect of rate ($F_{[2,15]} = 5.84; p = 0.024$). Post-hoc Bonferroni tests ($p < 0.05$) showed that DLs were significantly higher only for a rate of 300 PPS compared to 100 PPS. That result is consistent with previous studies that show that DLs generally deteriorate as rates approach or exceed 300 PPS (§2.3.2) and can be explained by saturation of neural discharge rates with increasing pulse rate that arises due to neural refraction (e.g., Shepherd and Javel, 1997; Miller et al., 2001; Shepherd et al., 2004).

Figure 3.2. Pulse rate DLs for experiment 1a plotted for each subject and for the subject group average. DLs are expressed in semitones and as a percentage of the reference pulse rate on the left and right vertical axes respectively. For individual subject data, the error bars indicate 95% confidence intervals for the estimated thresholds. For group average data, error bars indicate 5% least significant difference (LSD) of means.

### 3.3.2 Exp. 1b: Pitch comparison as a function of electrode place differences

#### 3.3.2.1 Methods

The pitch of two unmodulated electrical pulse trains that differed in site of stimulation was compared (ranked) in a 2I-2AFC procedure using a method of constant stimuli. The pulse rate of the reference stimulus was always 100 PPS. The rates of the probe stimuli were chosen to bracket the pitch elicited by the reference stimulus, as determined prior to the pitch ranking test. For example, if probe rates of 80 and 160 PPS elicited pitches
that were respectively clearly lower and higher than that of the reference, then probe stimuli of 80, 100, 120, 140, and 160 PPS were used in the pitch ranking test. Within each test, the probe was always presented on electrode 18 and the reference at a fixed electrode separation from the probe. A control case in which both the probe and reference stimuli were presented on the same electrode was also tested. Separations of ±1 electrode were examined in all subjects. Subjects who showed little effect of that separation were additionally tested with a larger separation of ±4 electrodes.

To reduce any confounding influence of loudness on pitch all stimuli were loudness matched to a 100 PPS pulse train presented on electrode 18 at C-Level, using the adaptive procedure described in Sec. 3.2.5. For each test condition, probe and reference pairs were presented ten times in randomised order within a run and at least two runs were conducted per condition. Subjects were asked to nominate which of the two stimuli was higher in pitch. No feedback was provided. For each reference electrode, a psychometric function was fitted using the bias-corrected percentage of responses for which the probe rate on electrode 18 was deemed higher in pitch than the reference. A criterion of $P_{c,max} = 50\%$ was used to determine the probe-rate that produced a pitch percept equivalent to that produced by the 100 PPS reference, henceforth referred to as the “equivalent-pitch rate” (EP-rate). The difference in semitones between the EP-rate and the reference rate is referred to as the “EP-rate offset”. Positive offsets indicate that the probe rate needed to be higher than the reference to produce equivalent pitch percepts. Error bars in subsequent figures of individual subject results plot 95% confidence intervals of the estimated EP-rate offsets so that variability in psychometric functions can be gauged.

3.3.2.2 Results and discussion

Equivalent-pitch rate offsets are plotted in Fig. 3.3 for each subject and for the subject group average as a function of electrode separation (where positive separations indicate more basal reference electrodes, e.g., +4 corresponds to electrode 14). The effect of electrode separation on EP-rate offset was examined in a two-way ANOVA. Note, S1 and S4 were not tested at separations of ±4 electrodes because they already demonstrated noticeable effects at the smaller separations of ±1 electrode and so those data points were missing from the analysis. A significant effect of electrode separation was observed ($F[4,16] = 5.42; p = 0.006$). Post-hoc Bonferroni tests showed that EP-rate offset for a separation of +4 electrode places (i.e., more basal) was significantly higher.
than those of all other electrode separations. No other significant effects were observed although a trend of negative EP-rates offsets for more apical electrode differences was apparent.

It is clear from Fig. 3.3 that results varied across listeners. Accordingly for each subject, 95% confidence intervals of the estimated EP-rates were used to compare the EP-rate offset for no electrode separation to those at the other separations. A significant effect of place differences was found at separations of ±1 electrode for S1 and S4, and for separations of -1, -4 and +4 electrodes for S2 and S3. Those effects were significant for more basal and apical electrode differences and suggest that the asymmetry in outcomes of the group analysis may have been due to the small sample size, subject differences, and missing data. In general, the results for S1-S4 are in agreement with those of (Pijl, 1997a; Zeng, 2002) which show that rate-pitch can be influenced by place of stimulation in a manner consistent with the general relationship between electrode position and place-pitch in CIs (e.g., Busby et al., 1994; Nelson et al., 1995). It is interesting however, that in the present study significant effects were observed at a reference rate of 100 PPS where the salience of rate-pitch is expected to be high. This was not generally the case for the low rate data of Pijl (1997a) and Zeng (2002) although results for some subjects in those studies support the present observations.

Figure 3.3. Equivalent-pitch (EP) rate offsets for experiment 1b plotted for each subject and for the subject group average as a function of electrode separation. Positive electrode separations indicate a more basal reference electrode site. EP-rate offsets are expressed in semitones and as a percentage of the reference rate of 100 Hz on the left and right vertical axes respectively. For individual subject data, the error bars indicate 95% confidence intervals for the estimated rate offset. For group average data, error bars indicate 5% LSD of means across subjects.

3 Presumably for these two subjects, separations of ±4 electrodes would also produce significant effects.
For S5 and S6, no significant effect of place differences was found, however it is possible that larger electrode separations than those tested may have produced a significant outcome. Because results for those two subjects differed from the others, they were further tested on several occasions, using different ranges of probe rates, and using an adaptive 2AFC procedure in which pulse rate was adjusted to determine an equivalent pitch. In all cases, there was no significant effect of place for these two subjects. Subject S6 was further tested at a higher rate of 300 PPS, for which rate-pitch is expected to be less salient than at lower rates. The EP-rate offset for an electrode separation of -4 places (i.e., electrode 22) was -2.43 semitones which was significantly different from zero (95% limits: -3.2 < μ < -1.4) indicating that the pitch of the 300 PPS pulse train on electrode 22 was perceived as lower than that on electrode 18. That outcome suggests that the variation in the effect of place across listeners is due to different weighting of rate and place contributions to pitch and that a stronger effect of place is expected when the rate cue is less salient.

### 3.3.3 Exp. 1c: Pitch comparison as a function of differences in stimulation level

#### 3.3.3.1 Methods

The pitch heights of unmodulated electrical pulse trains that differed in stimulation level (and hence loudness) were compared. All stimuli were presented to the same single site (electrode 18). The reference stimulus was always presented at a fixed pulse rate of 100 PPS, and at a fixed stimulation level that was higher or lower than the initial probe level. A control case in which both the probe and reference were presented at the same level was also tested. As in the previous experiment, the rates of the probe stimuli were predetermined to bracket the pitch elicited by the reference. The loudness of each probe stimulus was matched to that of a 100 PPS pulse train presented on electrode 18 at C-Level. The maximum level differences between the probe and reference stimuli that were examined varied across subjects and averaged from -7.5 (σ 2.6) to +14.5 (σ 10.9) CU, which spanned an average range of approximately -19.1% (σ 7.9%) to +32.1% (σ 16.3%) of each subjects’ EDR in CU. The same pitch ranking procedure as described in experiment 1b was used to determine EP-rate values.
Chapter 3: Pitch and Loudness of Amplitude Modulated Stimuli

3.3.3.2 Results and discussion

EP-rate offsets as a function of the stimulation level difference are plotted in Fig. 3.4 for each subject and for the subject group as a scatter plot. The stimulation level difference is calculated as the difference in CU between probe and reference and is expressed as a percentage of the EDR for a 100 PPS pulse train on electrode 18. The effect of stimulation level difference was examined using linear regression analysis. Results showed a significant increase in EP-rate as level increased ($slope = 0.095$, $r^2 = 0.42$, $p < 0.001$) although only 42% of the variance in the data was accounted for by that relationship. In separate analyses performed for each subject, significant effects of level differences were only observed for S2 ($slope = 0.279$, $r^2 = 0.95$, $p = 0.005$), S3 ($slope = 0.080$, $r^2 = 0.90$, $p = 0.014$) and S4 ($slope = 0.077$, $r^2 = 0.98$, $p = 0.009$). For those subjects, results may have been due to an apical shift in the low-frequency edge of the spatial excitation pattern with increasing stimulation level. That would be consistent with outcomes reported by Pijl (1997a) although those outcomes were based mainly on data at higher pulse rates than those used in the present study and although not tested, it is thus possible that relative pitch changes with level might be even larger at higher rates than those seen at 100 PPS. Results for subjects S1, S5, and S6 showed no significant effect of level on pitch. However, a trend of increasing pitch with increasing level was observed in results for S6 and S1 which is consistent with some of the data reported by Arnoldner et al., (2008) and Carlyon et al., (2010a). As per the previous experiment these differences in outcomes across subjects and studies might be accounted for by differences in weighting of rate and place cues to pitch across subjects (McKay et al., 2000).

Figure 3.4. EP-rate offsets for experiment 1c plotted for each subject and for the subject group average as function of the stimulation level difference. For the group average data, a line of best fit is plotted. Level differences (probe – reference level) in clinical units are expressed as a percentage of the electrical
3.4 Exp. 2: Pitch and loudness of modulated electrical stimuli

The pitch and loudness of amplitude-modulated high-rate pulse trains were compared to that of unmodulated low-rate pulse trains for which it has been shown that accurate identification of pitch based on pulse rate can be obtained (§2.3.2.3). The effects of modulation rate, depth, and shape were examined for stimuli presented at a comfortable loudness level, and again for a subset of the modulated stimuli at a lower presentation level. In addition to providing information about the relative pitch difference between modulated and unmodulated pulse trains, the stimulation levels needed to balance the loudness of those stimuli were also obtained. Outcomes from these tests can potentially direct choice of modulation functions and rules that reduce pitch differences between amplitude-modulated and unmodulated stimuli with the aim of providing accurate encoding of F0 rate information in a sound coding strategy.

3.4.1 Exp. 2a: Effects of modulation rate, depth and shape on pitch and loudness

3.4.1.1 Methods

Modulation rates of 100, 200, and 300 Hz, modulation depths of 12.5 or 25, 50, and 100% of the EDR (12.5% for 100 and 200 Hz and 25% for 300 Hz) and two modulation shapes were examined. Those shapes were sinusoidal amplitude modulation (SAM) of current and sharp-onset, exponential decay modulation (EDM) similar to that employed in the eTone strategy (see chapter 4). The EDM function examined in the present study consisted of a unit step function with an exponential decay to 90% of the modulation depth within the first 25% of the modulation period, see Fig. 3.5. This modulation function differed slightly from that described in chapter 4 which decayed exponentially in clinical units (i.e., log current) rather than current. However, the resulting difference is small over the electrical dynamic range of subjects in the present study and for CI subjects in general. For tests using a modulation frequency of 300 Hz, the shallowest modulation depth was increased from 12.5% to 25% of the EDR because it was found that many subjects had difficulty ranking pitch of those shallower stimuli consistently. As a consequence, data for a 12.5% modulation depth at a modulation frequency of 300 Hz could not be included in the analysis of variance.
Chapter 3: Pitch and Loudness of Amplitude Modulated Stimuli

All stimuli were loudness balanced using the procedure described in section 3.2.5. The unmodulated stimuli were presented at a comfortable loudness level as determined from subjects’ C-Levels at pulse rates of 100, 200, and 300 PPS. For each of those three nominal pulse rates, unmodulated stimuli of matched loudness were determined for pulse rates spanning approximately one octave around the nominal rate in rate steps of around 5-10%. For modulated stimuli, at each fixed modulation rate, depth, and shape, the peak level of the modulated pulse train was adjusted so that its loudness was matched to that of the unmodulated stimulus that had a pulse rate equal to the modulation rate. All pulses of the modulated stimulus were adjusted by the same number of clinical units, even if they were reduced below T-Level\(^4\), so that the modulation depth expressed as a function of the EDR in CU was maintained.

The same procedure as described in experiment 1b and 1c was employed to compare the pitch of the modulated and unmodulated stimuli and to determine their EP-rates. For each combination of modulation rate, depth, and shape, a fixed reference stimulus was generated. The probe stimuli consisted of the loudness matched unmodulated pulse trains needed to bracket the pitch elicited by the modulated stimuli, as determined prior to the experiment through informal pitch comparisons.

![Figure 3.5. Depiction of EDM stimulus plotted as current versus time.](image)

### 3.4.1.2 Results and discussion

**Pitch comparisons**

The EP-rate offsets are plotted in Fig. 3.6 as a function of modulation depth for each modulation shape. Results are shown for modulation rates of 100, 200, and 300 Hz in panels (a), (b), and (c) respectively. Those data reflect the perceived increase in pitch of

---

\(^4\) Only stimuli for conditions of 100% modulation depth contained sub threshold pulses. For experiment 2a, those conditions were few, and can be determined from EL-levels below 0% in Fig. 3.7. For experiment 2b, all stimuli with 100% modulation depth contained sub threshold pulses.
a modulated stimulus relative to an unmodulated pulse train stimulated at the modulation rate. In general, the results show that for low modulation rates and depths, the pitch of the modulated stimulus was substantially higher (e.g., for the 100 Hz SAM stimulus at a modulation depth of 12.5%, the EP-rate was 7 semitones, or 50%, higher) than that of the corresponding unmodulated pulse train, but the two converged as modulation depth was increased. In addition, EP-rate offsets decreased with increasing modulation rate (although at the highest modulation rate and shallowest modulation depth this could not be assessed due to missing data). However, because the effect of rate on the pitch of the unmodulated pulse trains can not be determined from the data, it is not clear how the absolute pitch of the modulated stimuli increase with increasing rate. Finally, the EP-rate offset for the EDM modulation shape was smaller than for SAM by approximately 1-2 semitones across all modulation rates and depths examined.

The effects of modulation rate (MR), depth (MD), and shape (MS) on EP-rate offset were evaluated using a four-way ANOVA (with subject as a random factor) which included all first and second order interactions. In that analysis, data were missing for all conditions with a 12.5% modulation depth combined with a 300 Hz modulation rate. In addition, for S1 data were missing for a modulation depth of 12.5%, and at the 300 Hz modulation rate, for which the subject had considerable difficulty performing the test. The analysis revealed strong effects of all three main factors: MS ($F_{[1,13]} = 67.4; p < 0.001$); MR ($F_{[2,13]} = 61.2; p < 0.001$); and MD ($F_{[2,13]} = 207.0; p < 0.001$). In addition, a significant interaction between MR and MD was observed ($F_{[3,13]} = 6.6; p = 0.008$). No other interactions were significant. The EP-rate offset for EDM ($\mu = 1.2$ semitones) was significantly lower than for SAM ($\mu = 2.6$ semitones). Mean EP-rate offset decreased with increasing modulation rate ($\mu = 3.2, 1.4, \text{and} 1.1$ semitones for MRs of 100, 200, and 300 Hz respectively). Post-hoc Bonferroni tests for the interaction between MR and MD showed that EP-rate offsets for MRs of 200 and 300 Hz were significantly lower than for 100 Hz at MDs of 12.5 and 50%. In contrast at a MD of 100%, only the offset for a MR of 300 Hz was significantly lower than for 100Hz. Mean EP-rate offsets also decreased with increasing modulation depth ($\mu = 4.2, 1.2, \text{and} 0.3$ semitones for 12.5, 50, and 100% MDs respectively). Post-hoc tests showed that EP-rate offsets were significantly lower at MDs of 50 and 100% compared to 12.5%. Although some of these results should be treated with caution as they are based on a data set that includes extrapolated data for MD = 12.5% at a MR = 300 Hz, they are consistent with the limited data gathered at a MD = 25%, and general trends in the data.
Figure 3.6. EP-rate offsets for experiment 2a plotted for modulation rates of 100, 200 and 300 Hz in panels (a), (b) and (c) respectively for each subject and for the subject group average as a function of modulation depth for each modulation shape. Modulation depth is expressed as a percentage of EDR for 1800 PPS pulse rate on electrode 18. Data for SAM modulation is plotted using filled symbols whereas unfilled symbols are used for EDM. Units for the vertical axes and the role of the error bars are the same as that used in Fig. 3.3.

Because of differences apparent in the data between subjects, the statistical analysis was repeated for individual subjects. Individual results for S1-S4 were consistent with most effects observed in the group analysis. Effects of MS were significant in data for S1 ($F[1,1] = 613.4; p = 0.026$), S3 ($F[1,3] = 26.0; p = 0.015$) and S4 ($F[1,3] = 16.7; p = 0.026$) in which EP-rate offsets for SAM were higher than those of EDM. A significant interaction between MD and MR was observed for S2 ($F[3,3] =$
Chapter 3: Pitch and Loudness of Amplitude Modulated Stimuli

44.6; \( p = 0.005 \) and S4 \( (F[3,3] = 14.6; \ p = 0.027) \) in which offsets were significantly lower for a MD of 12.5% compared to 100% for a MR of 100 Hz only. Significant effects of MR were observed for S1 \( (F[1,1] = 316.9; \ p = 0.036) \), S2 \( (F[2,3] = 320.5; \ p < 0.001) \) and S4 \( (F[2,3] = 119.8; \ p = 0.001) \) in which EP-rate offsets for 200 and 300 Hz were lower than those for 100 Hz. Effects of MD were significant in S2 \( (F[2,3] = 42.5; \ p = 0.006) \), S3 \( (F[2,3] = 113.9; \ p = 0.001) \) and S4 \( (F[2,3] = 112.7; \ p = 0.002) \) with lower offsets at 100% compared to 12.5%.

A very different outcome was found for S5-S6 who generally showed close pitch matches between modulated and unmodulated stimuli, irrespective of MD and MS. No significant effects of main factors or interactions were observed with the exception of MR \( (F[2,3] = 22.5; \ p = 0.016) \) for S6. Post-hoc Bonferroni tests showed that the EP-rate offset for 100 Hz was significantly lower than those at 200 and 300 Hz but mean differences were small (-1.72, -0.70 and -0.44 semitones respectively) and the effect was opposite to that seen in the group analysis. Subjects S5 and S6 both showed good ability to use pulse rate alone when judging the pitch of unmodulated stimuli in experiment 1b and 1c. In addition, their rate DLs in experiment 1a were amongst the lowest of all subjects tested. Thus, their small EP-rate offsets in the present experiment can not be attributed to poor pitch discrimination and an increase in those offsets would still be expected at some critically shallow modulation depth for which the stimuli would become indistinguishable from a high-rate unmodulated pulse train.

The results for S6 also differed from the other subjects in that EP-rate offsets at the lowest modulation rate were less than zero, implying that the pitch of the modulated stimulus was lower than that of the unmodulated pulse train. Similarly at a modulation rate of 300 Hz and a depth of 25% the EP-rate of the modulated signal was as much as 4.5 semitones lower than the unmodulated pulse rate and below that of all other modulation conditions tested. One possible explanation for those observations is that spike-rate adaptation (Zhang et al., 2007) and/or accommodation to pulse rate (Sly et al., 2007) was greater for the high pulse rate modulated stimuli (1800 PPS) than for the low-rate unmodulated pulse trains, which may have reduced average neural discharge rates and lowered the elicited pitch. This is discussed further in section 3.6.4 and evidence of greater adaptation/accommodation in the modulated stimuli for subject S6 (and to a lesser extent S5) is presented in the next section. Like S6, the results for S2 exhibit some cases (for modulation rates of 200 and 300 Hz) in which the pitch heights of the modulated stimuli were perceived lower than the unmodulated pulse trains. In those
cases however, there is no evidence of adaptation/accommodation in their data (see next section) and it is likely that other factors are responsible for results, see section 3.4.3.

Loudness matching

Figure 3.7 shows the level of the high stimulation rate modulated stimuli and the unmodulated stimuli with pulse rate equal to the modulation rate, that produce equal loudness, hereafter referred to as the “equivalent-loudness level” (EL-level). The EL-levels are plotted relative to the level of the unmodulated 1800 PPS pulse train on electrode 18 expressed as percentage of the EDR so that outcomes can be used directly when coding these types of signals in a sound processor strategy. Results are plotted for each modulation shape as a function of modulation depth for modulation rates of 100, 200, and 300 Hz in panels (a), (b) and (c) respectively. Also plotted are the EL-levels for the low-rate unmodulated stimulus (shown by the dashed line). In general, EL-levels for both modulation shapes increased with increasing modulation depth although little difference was observed between modulation depths of 50 and 100% of the EDR. EL-levels for EDM modulation were higher than those for SAM and converged with increasing modulation depth towards the level of the equally-loud unmodulated stimulus. As modulation depth was decreased, EL-levels for both modulation shapes approached zero. Increasing modulation rate caused EL-levels for both modulation shapes, as well as the difference between the two to decrease.

As per the analysis performed on EP-rate offsets, the effects of MR, MD, and MS on EL-levels were evaluated using a four-way ANOVA with subject as a random factor. Strong effects of all main factors: MR ($F[2,13] = 194.7; p < 0.001$); MD ($F[2,13] = 111.6; p < 0.001$); and MS ($F[1,13] = 152.9; p < 0.001$) were observed. In addition, significant interactions between MS and MD ($F[2,13] = 14.7; p < 0.001$), MS and MR ($F[2,13] = 77.9; p < 0.001$), MD and MR ($F[3,13] = 6.6; p = 0.006$), and all three main factors ($F[3,13] = 3.6; p = 0.042$) were observed. Post-hoc Bonferroni tests for the three-way interaction showed that EL-levels for EDM were significantly higher by approximately 10% EDR than those for SAM for all modulation depths but only for a modulation rate of 100 Hz. In addition, within modulation shape, no significant effect of MD or MR was observed for SAM, whereas for EDM, EL-levels were significantly higher for MDs of 50 and 100% compared to 12.5% but only at a modulation rate of 100 Hz, and for a modulation rate of 100 Hz compared to 200 and 300 Hz at all modulation depths.
Figure 3.7. Equivalent-loudness (EL) levels for experiment 2a plotted for modulation rates of 100, 200 and 300 Hz in panels (a), (b) and (c) respectively for each subject and for the subject group average as a function of modulation depth for each modulation shape. EL-levels are relative to the level of an unmodulated 1800 PPS pulse train on electrode 18 at the comfortable-loudness level and are expressed as % EDR. Data for SAM modulation is plotted using filled circles whereas unfilled circles are used for EDM. The dashed lines indicate the EL-level of the unmodulated low-rate stimuli. Units for the horizontal axes are the same as that used in Fig. 3.6. The error bars for individual subject data indicate standard errors of the means across test repeats whereas for group subject data they indicate 5% LSD of means across subjects.

Trends in most subject’s EL-levels were consistent with effects of modulation rate, depth and shape observed in the group analysis, with the exception of S1 and S4. For S1
Chapter 3: Pitch and Loudness of Amplitude Modulated Stimuli

there is incomplete data and so an adequate comparison to group results is not possible. For S4, measurement of levels near the comfortable loudness level were compromised by non-auditory facial nerve stimulation and so interpretation of those data which show very little effect of modulation depth and rate on loudness is problematic.

For most subjects and conditions, EL-levels were less than or equal to those of the unmodulated pulse trains (depicted by the dashed lines in Fig. 3.7). However, for S5 and S6, it is interesting to observe that the stimulation level of the high pulse-rate modulated stimulus was in some cases greater than that of an equally-loud low-rate unmodulated pulse train (e.g., for S5: EDM at the highest MD and MRs of 100 and 200 Hz; and for S6: EDM at MDs of 50 and 100% and a MR of 100 Hz). Those data for S5 and S6 can most likely be attributed to greater spike-rate adaptation and/or accommodation to the high pulse rate of the modulated stimuli compared to the low-rate unmodulated stimuli.

3.4.2 Exp. 2b: Effects of presentation level on pitch and loudness

3.4.2.1 Methods

Experiment 2a was repeated for a modulation rate of 100 Hz only, using a lowered “mid-loudness” level that resulted in approximately halved loudness. That level was determined by asking the subject to rate the loudness of the unmodulated stimuli on a scale of 1 to 10 (just audible to maximum loudness before discomfort) and by reducing the stimulation level to the point where the loudness rating was reported to be approximately half that of the C-Level stimulus. This test served to determine whether the effects of modulation depth and shape on pitch and loudness observed in experiment 2a were independent of presentation level.

3.4.2.2 Results and discussion

Pitch matching

EP-rate offsets for the 100 Hz modulated stimuli presented at a mid-loudness level are plotted in Fig. 3.8 for each modulation shape as a function of modulation depth. The effects of modulation depth, shape, and presentation level, on EP-rate offset were evaluated using a four-way ANOVA which included the 100 Hz modulation data collected at C-Level in experiment 2a so that the effects of presentation level could be determined. In that analysis, data was not available for S1 at 12.5% MD and for S5 and
Chapter 3: Pitch and Loudness of Amplitude Modulated Stimuli

S6 at 100% MD. Significant effects of all main factors: MD ($F[2,7] = 138.0; p < 0.001$); MS ($F[1,7] = 33.1; p < 0.001$), and presentation level ($F[1,7] = 20.7; p = 0.003$); and the interactions between presentation level and MD ($F[2,7] = 17.2; p = 0.002$) and presentation level and MS ($F[1,7] = 6.84; p = 0.035$) were observed. Post-hoc tests for the interaction between presentation level and MD showed that for stimuli presented at the comfortable loudness level (C-Level) EP-rate offsets were significantly lower at MDs of 50 and 100% than those at 12.5%. In contrast for stimuli presented at the mid-loudness level (Mid-Level) this was only the case for a MD of 100%. For the interaction between presentation level and MS, post-hoc tests showed that EP-rate offsets for EDM were significantly lower than those for SAM at C-Level but not at Mid-Level. In addition, for SAM at a MD of 12.5%, EP-rate offsets were significantly lower at Mid-level than at C-Level.

In general, EP-rate offsets at Mid-Level were lower than those presented at C-Level, particularly for the shallower modulation depths and for SAM stimuli. The differences in EP-rates between SAM and EDM and across modulation depths were much smaller at Mid-Level than at C-Level. Those outcomes demonstrate that while deep modulation is needed to provide a good pitch match to a low-rate unmodulated pulse train at comfortable levels, the requirement diminishes at lower levels.

Figure 3.8. EP-rate offsets for experiment 2b for stimuli presented at a mid-loudness level and a modulation rate of 100 Hz are plotted for each subject and for the subject group average as a function of modulation depth for each modulation shape. Data symbols, units for the horizontal and vertical axes, and the role of the error bars are the same as that used in Fig. 3.6.
Stimulation level differences between C- and Mid-Level

Averaged across the six subjects, the difference between C-Level and Mid-Level at 100 PPS was 15.0 (± 11.1) CU, or 31.4% (± 11.1%) of the EDR at 100 PPS. The average difference between the C- and Mid-Level loudness-matched 1800 PPS unmodulated stimuli was 14.6 (± 7.9) CU, or 23.7% (± 14.0%) of the EDR at 1800 PPS.

Loudness matching

Equivalent-loudness levels for stimuli presented at a level approximately half as loud as those used in experiment 2a are plotted in Fig. 3.9 for a modulation rate of 100 Hz. The effects of modulation depth and shape, and presentation level on EL-levels were evaluated using a four-way ANOVA which included the 100 Hz modulation data collected at C-Level in experiment 2a. Significant effects of MS ($F[1,7] = 126.0; p < 0.001$), MD ($F[2,7] = 34.2; p < 0.001$); and the interaction between MS and MD ($F[2,7] = 9.74; p = 0.009$) were observed. Post-hoc tests for the interaction showed that EL-levels for SAM were significantly lower than those for EDM at MDs of 50 and 100%. For SAM, no effect of MD was observed whereas for EDM, EL-levels were significantly lower at a MD of 12.5% compared to 50 and 100%. More generally however, the effects of modulation depth and shape on EL-levels were similar for C- and Mid-Levels (as evidenced by the lack of interaction between presentation level and those factors). That outcome differs from that seen for EP-rates which showed decreasing effects of modulation depth and shape with decreasing presentation level.

None of the individual subject EL-levels for the modulated stimuli was greater than those of the unmodulated 100 PPS pulse trains at Mid-Level presentation, which is in contrast to some of C-Level results for S5 and S6 in experiment 2a. That implies that if those C-Level results are due to adaptation/accommodation to pulse rate, that mechanism is less active at the lower Mid-Level in those two subjects.
Chapter 3: Pitch and Loudness of Amplitude Modulated Stimuli

3.4.3 Analysis of effects of stimulation level differences on pitch matching

Because changes in stimulation level can affect rate-pitch (e.g., Pijl 1997a, Arnoldner et al., 2008; Carlyon et al., 2010a) pitch differences observed between stimuli in the present experiments might be attributable to differences in stimulation level amongst stimuli. In experiment 1c, a significant effect of stimulation level on the pitch of 100 PPS pulse trains was noted for subjects S2, S3, and S4. For S2, results of experiment 2a showed an increase in EL-level of approximately 13.5% of the EDR for the 100 Hz EDM stimulus when modulation depth was increased from 12.5 to 100%. According to that subject’s results in experiment 1c, this would correspond to a decrease of approximately 5 semitones in EP-rate, which is close to but less than the 8.8 semitone decrease observed in experiment 2a. It is thus possible that effects of stimulation level on the apical edge of excitation may in part account for that subject’s pitch result (including the data across modulation rates where that subject’s EP-rate offsets fell below zero at higher rates). In contrast for subjects S3 and S4, examination of the EP-rate and EL-level data across modulation depth, rate, and shape revealed many instances where substantial changes in EP-rate occurred with relatively little change in EL-level. For instance for S3, EP-rate for SAM stimuli at 100 Hz dropped by approximately 12 semitones for an increase in modulation depth of 12.5 to 100% but the EL-level only changed by 2.8% of the EDR. Similar effects were observed in data for S4. Subjects S1, S5 and S6, showed no significant effects of stimulation level on pitch in experiment 1c.
These results show that for five of the six subjects changes in pitch in the present experiment can most likely be attributed to factors other than stimulation level differences.

To further examine the effects of stimulation level on pitch, EP-rate offsets (relative to unmodulated pulse rate) were compared to EL-levels (relative to unmodulated pulse train level) using a linear regression analysis applied to the data from experiment 2a. When data from all subjects and conditions were included, a significant negative correlation was observed ($slope = -0.639$, $r^2 = 0.18$, $p < 0.001$) which suggests that a shift in the low-frequency edge of the excitation pattern to changes in stimulation level may account for some of the pitch results. However, only a small proportion of the variance in the data (18%) was accounted for, and in a similar analysis of data from experiment 2b, no significant correlation was observed. Taken together the results suggest that the pitch results were mainly attributable to factors other than stimulation level in the majority of subjects. Nevertheless, the possibility that some subjects use a weighted combination of rate and place of stimulation to judge pitch in some circumstances can not be ruled out (McKay et al., 2000).

Additionally, the above analyses do not take into account changes in absolute pitch with increasing pulse rate. It is possible that the reduction in levels as pulse rate increases that were needed to maintain equal loudness may have limited growth in average neural discharge rates and therefore also pitch height. That may also in part account for the reduction in EP-rate offsets with increases in pulse/modulation rate in experiment 2a. Such a mechanism is consistent with the behavioural data reported by Carlyon et al., (2010a) in which rate-pitch was shown to decrease with decreasing level.

### 3.5 Exp. 3: Pitch of modulated acoustic and unmodulated electrical stimuli

The pitch of low-rate unmodulated electrical pulse trains was compared to that of complex harmonic tones presented acoustically for three of the six CI subjects in the present study who used a contralateral hearing aid. That comparison aimed to determine the absolute relationship between electrical rate-pitch and acoustic pitch. While previous studies (§2.3.2.3) have shown that relative changes in electrical pulse rate can produce pitch intervals similar to those elicited in acoustic hearing to changes in F0, it is not clear whether absolute pitch is similar for electric and acoustic rates. Blamey et al., (1996)
Chapter 3: Pitch and Loudness of Amplitude Modulated Stimuli

compared the pitch of unmodulated electric pulse trains to acoustic pure tones in CI users with some residual hearing in their non-implanted ear. On average pitch was shown to increase more rapidly as a function of pulse rate than pure tone frequency. As noted by the authors, the data could be explained by increased influence of electrode place and decreased rate-pitch salience as pulse rate increases (e.g., Pijl, 1997a; Zeng, 2002). In an effort to deter subjects from judging pitch on the basis of place information in the present experiment, low-pass filtered acoustic complex harmonic tones were employed rather than pure tones. The harmonic tones produce multiple resolved places of spatial activation along the cochlea in contrast to a single site for the electrical stimuli. That difference in spatial activation pattern was intended to reduce the likelihood that pitch would be compared on the basis of a place. In contrast, an acoustic pure tone produces a single region of activation that is more similar to that of an electrical pulse train.

3.5.1 Exp. 3a: Effects of acoustic modulation rate

3.5.1.1 Methods

The pitch heights of complex harmonic tones presented acoustically to a hearing aid (HA) located contralateral to the CI ear, were compared to those of unmodulated pulse trains presented electrically. Participants comprised a subset of the six subjects (S3, S4, and S6) that wore a hearing aid. Details of their hearing aids and aided thresholds are presented in Table 3.2.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Hearing Aid</th>
<th>125</th>
<th>250</th>
<th>500</th>
<th>1K</th>
<th>1.5K</th>
<th>2K</th>
<th>3K</th>
<th>4K</th>
</tr>
</thead>
<tbody>
<tr>
<td>S3</td>
<td>Phonak Claro 22</td>
<td>35</td>
<td>35</td>
<td>50</td>
<td>85</td>
<td>85</td>
<td>85</td>
<td>90</td>
<td>90</td>
</tr>
<tr>
<td>S4</td>
<td>Siemens CIELO 2P</td>
<td>20</td>
<td>15</td>
<td>35</td>
<td>35</td>
<td>50</td>
<td>55</td>
<td>65</td>
<td>90</td>
</tr>
<tr>
<td>S6</td>
<td>Siemens Nitro 300SP</td>
<td>70</td>
<td>65</td>
<td>65</td>
<td>55</td>
<td>40</td>
<td>40</td>
<td>45</td>
<td>35</td>
</tr>
</tbody>
</table>

Table 3.2. Aided warble tone thresholds in dB hearing level. Tone duration = 500 ms. S3 and S4 both had steeply sloping high frequency losses whereas S6 had more low frequency loss than at higher frequencies.

EP-rate offsets for unmodulated electrical stimuli presented at the same electrode used in previous experiments (electrode 18) were measured for contralateral acoustic fundamental frequencies (modulation rates) of 100, 200, and 300 Hz. The same methodology used in previous pitch ranking experiments was employed but in this case
the reference was the acoustic harmonic complex described in section 3.2.3. The EP-rate offset in this case shows the increase in electrical rate needed to match the pitch elicited by the acoustic signal. The electrical stimuli were presented at a comfortable loudness (i.e., at the same levels as used in experiment 2a) and the acoustic stimuli were loudness matched to them using the procedure described in section 3.2.5.

3.5.1.2 Results and discussion

Figure 3.10 shows EP-rate offsets for the unmodulated electrical stimuli compared to the acoustic stimuli for F0s of 100, 200, and 300 Hz. The effects of F0 on EP-rate offset were evaluated using a two-way ANOVA with subject as a random factor. No significant effect of F0 was observed and the mean EP-rate offsets were 0.16, 0.25, and 1.88 semitones at 100, 200, and 300 Hz respectively. In addition, the 95% confidence limits (least significant difference of means) depicted by the error bars in the group average results show that EP-rate offsets were not significantly different from zero. The group average data show a good match (i.e., within a few semitones) between the pitch of the electrical pulse trains and the acoustic tones. However, some subject variability was observed mainly for S3 at 300 Hz where the pitch-matched electrical pulse rate was approximately 5 semitones higher than F0. The resulting trend in the group average data showing larger EP-rate offsets at 300 Hz suggests a decrease in electrical pitch relative to that of acoustic stimulus at that rate, which might be significant in a larger subject pool. That outcome would be consistent with saturation of the rate-pitch percept with increasing pulse rate.

Figure 3.10. EP-rate offsets for experiment 3a and 3b plotted for subjects S3, S4 and S6 and their group average as a function of acoustic F0 and electrode place for the lowest F0. Data for stimuli presented on electrode 18 are plotted by filled circles. Unfilled, or partly-filled, symbols are used for stimuli presented on electrodes more basal or apical to El 18. Units for the vertical axes and the role of the error bars are the same as that used in Fig. 3.6.
Chapter 3: Pitch and Loudness of Amplitude Modulated Stimuli

3.5.2 Exp. 3b: Effects of electrical place of stimulation.

3.5.2.1 Methods

Using the same methodology as described in experiment 3a, EP-rates for unmodulated electrical stimuli presented at two additional electrode places either side of electrode 18 were measured for an acoustic F0 of 100 Hz. For subjects, S3 and S6, the two electrodes activated were 14 and 22 (i.e., ±4 electrodes) whereas for S4, electrodes 17 and 19 were used (±1 electrode). Those electrode places correspond to the most distant electrode sites examined in experiment 1b.

3.5.2.2 Results and discussion

EP-rate offsets for electrical stimuli presented at neighbouring electrode places to that used in experiment 3a are plotted in Fig. 3.10 (unfilled symbols). The effect of electrode place (including data for stimuli presented on electrode 18 at 100 Hz) was evaluated using a two-way ANOVA (with subject as a random factor). No significant effect of electrode place was observed and EP-rate offsets for stimuli presented at electrode 18, at a more apical site, and at a more basal site were 0.16, -0.86, and -2.23 semitones respectively. According to the 95% confidence interval, the EP-rate offset for stimuli presented more apical to electrode 18 was not significantly different to zero. For stimuli presented at a more basal electrode, the EP-rate offset differed by more than one standard deviation compared to electrode 18 (mainly due to results of S3) implying that in a larger subject pool, the pitch percepts elicited by those electrical stimuli may be demonstrably higher than for the acoustic stimuli. That would be consistent with the relationship between electrode position and pitch in electrical stimulation, where more basal sites generally elicit a higher pitch, as well as the results of experiment 1b. Of the three subjects tested in the present experiment, only S3 and S4 showed a significant effect of place on EP-rate in experiment 1b. However, in the present experiment both subjects showed smaller or conflicting effects of place on pitch. Possible reasons for that outcome include the dissimilarity in spatial patterns of excitation to acoustic and electric stimuli, a change in “matching” criteria in bimodal stimulation compared to electric alone, and the greater familiarity with temporal information presented on electrode 18 compared to other electrodes by the time experiment 3 was conducted.

Although the data are preliminary, and subjects had impaired auditory hearing, the results suggest good correspondence between the pitch of low-rate electrical pulse trains
Chapter 3: Pitch and Loudness of Amplitude Modulated Stimuli

and acoustic complex harmonic tones matched in F0 rate, even when no attempt is made to match mean place of activation across ears. A possible alternative interpretation of the data is that subjects ignored the acoustic stimuli altogether in this experiment and based responses on whether the electrical pulse rate was toward the top or bottom of the range of rates presented in a block of trials (e.g., Carlyon et al., 2010b). However, for all three subjects the method employed to determine the range of pulse rates to be used for each condition required the listener to directly compare the pitch of acoustic and electric stimuli, such that clearly higher and lower pitch percepts were heard for the electric stimulus. In addition extended testing with subject S6, showed no significant differences in EP-rate offsets when the range of probe pulse rates used in the test was altered.

3.6 General Discussion

3.6.1 General outcomes

In experiment 1a, the ability to discriminate the pitch of unmodulated pulse trains on the basis of rate was shown to vary across listeners, as is consistent with previously reported data. Excluding one poor performer, group averaged rate-pitch DLs at low rates were similar to those reported in a review by Moore and Carlyon (2005), and deteriorated as rates approached 300 PPS (§2.3.2). The pitch heights of unmodulated pulse trains were compared to one another when differences in place, or level, were introduced in experiments 1b and 1c, respectively. Pitch was influenced by variations in place (§2.3.2.5; §2.3.2.7) consistent with a tonotopic representation of pitch (§2.3.1) for four of the six subjects, but not the other two. Stimulation level influenced pitch in a manner consistent with an apical shift of the low-frequency edge of the excitation pattern with increasing level (Pijl, 1997a) for three of the subjects. For the other three, pitch remained relatively fixed with changes in stimulation level, although an opposite trend of increasing pitch with level was noted for two of those subjects, which may be attributed to a temporal pitch coding mechanism (e.g., Carlyon et al., 2010a).

In experiment 2, the pitch heights of amplitude-modulated pulse trains with a high pulse rate were compared to those of unmodulated low-rate pulse trains. To address the four research questions described in the introduction, the results demonstrated that: (1) for low modulation-rates, amplitude-modulated pulse trains generally elicit a pitch that is higher than an unmodulated stimulus that has a pulse rate equal to the modulation rate, but the two converge as modulation depth is increased. That result is consistent with the
limited data available from McKay et al. (1995). For SAM stimuli at low modulation rates and comfortable loudness, modulation depths close to 100% of subjects’ EDR were generally needed to produce similar pitch heights to those produced by the unmodulated pulse trains; (2) application of deep modulation while holding peak stimulation level fixed can reduce loudness by a significant portion of a subject’s EDR. This was particularly evident at low modulation rates for EDM and will be discussed with relevance to F0 coding strategies in the next section; (3) as modulation rate increased to 200-300 Hz, equivalent-pitch rate offsets decreased towards zero. That is as expected if pitch percepts elicited by both modulated and unmodulated stimuli become saturated or less salient at high rates, as implied by previous studies (e.g., Shannon 1983, 1992; Tong and Clark, 1985; McKay et al., 1995; Zeng, 2002). In addition, results showed that pitch differences between modulated and unmodulated stimuli were smaller at the mid-loudness level than at the comfortable loudness level. Those two outcomes indicate that the need to apply deep modulation in a sound coding strategy is reduced at higher modulation rates and lower presentation levels; and (4) the pitches elicited by the EDM stimuli were generally lower than those elicited by SAM stimuli for equal modulation depths. For a modulation depth of 50% (or greater) EDM elicited pitch heights that were similar to those of unmodulated pulse trains. That demonstrates an improved ability to encode F0 compared to SAM, which required 100% modulation depth to achieve the same outcome.

In experiment 3, the pitch heights of low-rate electrical pulse trains were compared to those of acoustic harmonic tones in three of the six subjects who wore a contralateral hearing aid. These preliminary results demonstrated a good correspondence between the pitch of the electrical pulse trains and the acoustic tones when pulse rate was equal to acoustic F0. That outcome is in agreement with earlier findings that show that musical intervals can be coded using electrical rate (§2.3.2.3), but further suggests an absolute correspondence between acoustic pitch and electrical rate-pitch over a limited range of conditions.

3.6.2 Relevance to sound coding strategies

The preliminary results of experiment 3 suggest that absolute pitch elicited by electric pulse rate can be very similar to that derived from F0 in acoustic hearing, although effects of place and level may also influence pitch, as was observed in experiments 1 (b) and (c). Together with outcomes from other studies (Pijl and Schwarz, 1995;
McDermott and McKay, 1997) which show that similar pitch intervals result from matched changes in pulse rate in CIs, and F0 in acoustic hearing, it can be implied that coding of F0 via rate can provide an accurate representation of F0 in CI hearing, at least for low F0s. In experiment 2, similar pitch percepts were elicited by unmodulated and amplitude-modulated pulse trains when sufficiently deep modulation was employed, so that accurate encoding of F0 is possible using modulated stimuli. Finally, results of experiment 1a (excluding S1) and other studies (Moore and Carlyon, 2005) show that average pulse rate discrimination thresholds of around 1 semitone (with a range of a 0.5-3 semitones) are possible, at least for low F0s. Combined these outcomes indicate that if F0 is well coded using rate or modulation rate, CI recipients may discriminate F0 with sufficient accuracy to hear the intervals that delineate Western musical scales. It should be noted that the described outcomes are for stimulation on single electrodes, so that applicability to sound coding systems assumes that similar behaviour is upheld when activating many electrodes in quick succession.

Studies examining pitch perception in CI recipients using existing clinical coding strategies have generally shown poor outcomes in musical pitch tests (e.g., Gfeller and Lansing, 1991; Gfeller et al., 2005; Looi et al., 2004, 2008). Most such strategies employ a bank of band-pass filters (BPFs) to estimate the narrow-band envelope of the signal in each channel. For harmonic signals, such as voiced vowels or tonal musical sounds, these envelope signals contain a varying degree of F0 amplitude modulation. For channels in which only one harmonic of F0 is present, the envelope signal will contain no modulation. For channels in which two or more F0 harmonics are present, beating between those harmonics will produce some degree of F0 modulation in the envelope. The depth and shape of modulation will be dependent on the relative amplitudes and phases of each harmonic within the channel. Under ideal conditions such as when two or more similar amplitude sine-phase harmonics are present in a channel, the resultant modulation will be sinusoidal and very deep. For cases in which harmonics differ substantially in amplitude and/or phase, the resultant modulation will become less sinusoidal and its depth may be reduced. In addition, modulation across channels may not necessarily be in phase resulting in a reduction in the effective modulation depth seen by neurons responsive to multiple electrodes (McKay and McDermott, 1996).

For voiced and tonal sounds encountered in everyday listening conditions, F0 modulation coded by existing clinical strategies is often sinusoidal-like in shape, but
Chapter 3: Pitch and Loudness of Amplitude Modulated Stimuli

relatively shallow and quite variable across different sounds. For instance, in an unpublished laboratory analysis by the present authors of sung vowels presented in the sound field to the ACE strategy, the effective modulation depth averaged across apical-to-mid channels was approximately 10-20% of the EDR. Accordingly, the results from experiments 2 and 3 suggest that many subjects using clinical strategies will perceive a substantially higher pitch for low F0s than in normal hearing, which vary across sound sources and acoustic environments that affect the shape and depth of F0 modulation. In addition, as F0 increases towards CI users’ upper limit of rate discrimination, rate-pitch may increase at a progressively shallower rate due to saturation or become difficult to discern due to reduced cue salience. The same effect can be seen in the data from experiment 2 with SAM stimuli with low modulation depths, which show a decrease in EP-rate offsets with increasing modulation rate. Furthermore, the reduced salience of F0 rate information due to relatively shallow modulation in clinical strategies means that the judgment of pitch is more likely to be influenced by confounding effects of place-coding.

Using present stimulation methods it appears that little can be done to overcome the upper limits of rate-pitch discrimination and thus it is ineffective to code F0s beyond approximately 300 or 400 Hz using rate or modulation rate information. However for lower F0s, the results of experiment 2 suggest that a salient and accurate representation of F0 can be achieved by encoding F0 information using SAM at modulation depths close to 100% EDR. Alternatively, sharp onset, rapid decay modulation, such as EDM, can also provide accurate encoding of F0 at a shallower modulation depth of 50% of the EDR, although changes in loudness may need to be compensated for when applying deep modulation to the stimulus envelope particularly for low F0s and for EDM. Assuming loudness is well coded and/or that subjects have become accustomed to the loudness coded by existing clinical strategies, such compensation should aim to maintain similar loudness when applying deep F0 modulation in rate-pitch coding strategies such as eTone (see chapter 4). If sinusoidal-like amplitude modulation encoded by clinical strategies for typical real-world sounds produces an effective modulation depth of approximately 10-20% of the EDR in general, then according to the results from experiment 2a, the stimulation level for an F0 of 100 Hz when using EDM with a modulation depth of 50% in eTone would need to be increased by approximately 10-15% of the EDR to maintain loudness (assuming that loudness is integrated similarly across channels to each form of modulation). That estimate is consistent with level
modifications needed to maintain loudness in a study comparing pitch and speech perception performance between ACE and eTone (see chapter 5, §5.2.3). However, the effect of modulation rate on EL-level in experiment 2a indicates that any level compensation applied by strategies needs to decrease with increasing F0. In addition, in experiment 2b it was shown that pitch differences between modulated and unmodulated stimuli decreased with decreasing presentation level and so the need to employ deep modulation might be relaxed at lower presentation levels. Thus loudness differences at lower presentation levels might instead be managed in eTone or similar strategies by decreasing the amount of F0 modulation applied.

3.6.3 A conceptual neural mechanism explaining temporal pitch

The outcomes of experiment 2 can be explained by considering the effects of pulse rate and level on neural recruitment. It is thought that for unmodulated low-rate pulse trains presented at a comfortably loud stimulation level, a high degree of entrainment occurs for those neurons that are stimulated at high levels relative to their thresholds (see §2.3.2.1). Neurons that are stimulated at relatively lower levels may discharge at a lower sub-multiple of the pulse rate. An increase in stimulation level can therefore promote greater overall entrainment to the pulse rate. As pulse rate is increased with a fixed stimulation level, entrainment decreases but overall neural discharge rates increase up to a saturated limit, due to refractory effects. If loudness, rather than level, is held fixed while pulse rate is increased, average discharge rate may not necessarily increase due to the reduction in level needed to maintain loudness at higher rates.

The effects of rate and level on discharge rates are similar to those seen in behavioural measures of rate-pitch. It is thus thought that rate-pitch is determined by some average of neural discharge rates (e.g., Carlyon et al., 2002; van Wieringen et al., 2003) although details are unclear, particularly at higher rates. For high pulse rate amplitude-modulated stimuli such as those used in the present study, neurons that are presented with levels well above threshold for all pulses in the modulation cycle are able to fire to any pulse, and therefore may be said to be responsive to the stimulation pulse rate, or more accurately to sub-multiples of the pulse rate. For neurons for which that is not the case, stimulation pulses in the trough of the modulation cycle may no longer elicit responses. The reduced period during which those units can fire may cause them to fire only once within a modulation cycle and they can therefore be considered to being
Chapter 3: Pitch and Loudness of Amplitude Modulated Stimuli

responsive only to the modulation rate\(^5\), or a sub-multiple thereof. Overall, the population response can potentially comprise a continuum of discharge rates ranging from the modulation rate (and submultiples thereof) to refractory limited sub-multiples of the pulse rate, which will be governed by the distribution of neural thresholds and variations in stimulation level. Within this context, each of the results from experiment 2 is considered in turn.

Modulation depth: As modulation depth increases, the relative proportion of neurons responding to the modulation rate as opposed to sub-multiples of the pulse rate will increase. The average discharge rate will decrease towards the modulation rate and the elicited pitch will decrease towards that of an unmodulated pulse train that has a pulse rate equal to the modulation rate. This is in agreement with the data shown in Figs. 3.6 and 3.8 for the modulated stimuli which show a decrease in EP-rate offsets to zero (i.e., to a pitch height similar to that of the unmodulated stimuli) with increasing modulation depth. Conversely when modulation depth decreases, the average discharge rate increases towards that produced by an unmodulated high-rate pulse train.

Modulation shape: For sufficiently deep modulation, the modulation peak for EDM is substantially narrower than that for SAM. Accordingly, the opportunity for neurons to fire more than once per modulation cycle is reduced more for EDM than SAM, and accordingly the rate-pitch will be lower. The lower EP-rate offsets for EDM compared to SAM stimuli in Figs. 3.6 and 3.8 support this conjecture.

Presentation level: At lower presentation levels, a greater portion of the responsive population is stimulated at sub-threshold levels during the trough of the modulation cycle. The relative proportion of neurons responding to the modulation rate is therefore increased and the elicited pitch shifts towards that produced by an unmodulated low-rate pulse train as was observed by the lower EP-rate offsets for Mid-Level presentation shown in Fig. 3.8 compared to those at C-Level in Fig. 3.6.

Modulation rate: If a substantial portion of the population is stimulated well above threshold, refractory effects may limit responses to both unmodulated and modulated stimuli, so that the pitch percepts converge with increasing rate. At lower levels, such as those needed to preserve matched loudness at high pulse or modulation rates, pitch percepts may be similarly limited, but more by reduced response probability than

\(^5\) For the case where pulse rate is not an integer multiple of modulation rate, average discharge rates over several cycles will encode the modulation rate.
refractory limitations. The decrease in EP-rate offsets towards zero with increasing modulation rate in Fig. 3.6 support these conjectures.

3.6.4 Differences between subjects

Some variability in results across subjects was observed in experiment 2 in which general outcomes could be attributed to results of S1-S4 but not to those of S5-S6. Pitch ranking results for those two subjects showed little difference in pitch between the modulated and unmodulated stimuli, no significant effect of modulation depth, and little effect of modulation rate. Those subjects’ results also differed from those of the other subjects in experiments 1b and 1c. It is possible that those subjects ignored the reference stimuli and judged pitch according to the range of probe rates presented within a block of trials (e.g., Carlyon et al., 2010b). However, repeated tests on different occasions using different ranges of probe rates, and using an adaptive 2AFC procedure, all confirmed the reported data and suggested that those subjects based their judgment of pitch mainly on temporal information in the electrical signals. For those two subjects, the loudness matched levels of the modulated stimuli were in some cases greater than the level of the unmodulated low-rate stimuli which was attributed to a greater degree of spike-rate adaptation (Zhang et al., 2007) and/or accommodation (Sly et al., 2007) to the high pulse rate in the modulated stimuli compared to the unmodulated low-rate stimuli\(^6\). Spike-rate adaptation causes a reduction in firing probability (due to ongoing spikes) and generally increases with increasing pulse rate. For modulated stimuli, its main effect is to reduce high pulse rate responses and shift the pitch towards the modulation rate. For low-rate unmodulated pulse trains, pitch is little affected. Consequently the difference in pitch between the modulated and unmodulated stimuli is reduced by spike-rate adaptation. Accommodation reduces neural firing probability when nerves are presented with pulses, but do not discharge, and potentially lowers the average discharge rate below that of the modulation rate. That might explain the results for subject S6 for whom the pitch of the modulated stimuli was ranked lower than that of the unmodulated pulse trains in some cases. Although spike-rate adaptation and/or accommodation may have contributed to the outcomes observed with subjects S5-S6, it is likely that other factors such as pathology, neural survival, and proximity of electrodes to neurons, also

\(^6\) It is possible that accommodation and/or adaptation to the high pulse rate did occur for subjects S1-S4, even though there was no direct evidence for it in their EL-level data. It is possible that the EL-levels for the high-rate modulated stimuli were lower than those of the low-rate unmodulated stimuli simply because of the reduction in levels needed to match loudness across pulse rates.
Chapter 3: Pitch and Loudness of Amplitude Modulated Stimuli

played a role. In addition, it may be possible that rate-pitch is derived from different weightings of the various neural discharge rates across listeners.

3.7 Conclusion

Average pulse rate discrimination thresholds for low-rate unmodulated electrical pulse trains obtained by subjects (excluding one who had difficulty with the test) were consistent with data reported in previous studies. At higher pulse rates, discrimination thresholds increased, consistent with discharge rate limitations imposed by the refractory behaviour of auditory neurons. For some of the subjects, both place and level of stimulation influenced the pitch height of unmodulated pulse trains, which can be accounted for by tonotopic place-pitch and the apical edge of the stimulus excitation field. However other subjects, and in particular those with the best pulse rate DLs, showed little influence of place or level on pitch.

For amplitude-modulated pulse trains and low modulation rates, results for four of the six subjects showed that SAM stimuli elicited a higher pitch than equally-loud unmodulated pulse trains that were matched in pulse rate to the modulation rate. However, the pitch of the two signals converged as modulation depth or rate was increased, or as presentation level was decreased. For equal modulation depths, the pitch elicited by EDM stimuli was generally found to be lower than for SAM stimuli. Those outcomes can be explained by the effects of pulse-level and -rate on neural responses, combined with the assumption that rate-pitch is based on a weighted average of neural discharge rates. For those few subjects who showed little effect of place or level differences on the pitch of unmodulated pulse trains, pitch differences between modulated and unmodulated stimuli were small.

The loudness of low modulation-rate pulse trains generally decreased with increasing modulation depth but little effect was seen for depths greater than approximately 50% of the EDR. The effects were greatest for EDM, and substantially higher stimulation levels were needed to match the loudness of stimuli across modulation depths than was the case for SAM. At higher modulation rates, the effects of modulation depth and shape on loudness generally decreased and were not significant at the highest rate examined of 300 Hz.

For those subjects with residual hearing in the non-implanted ear, a good correspondence between the pitch of electrical pulse trains and acoustic harmonic tones was observed when pulse rate was similar to F0. Even for the highest rate of 300 Hz, the
pitch of the acoustic and electric stimuli were well matched, although the pitch of the unmodulated pulse train tended to be lower than the acoustic stimulus, presumably due to the effects of neural refraction. Although preliminary, those data suggest an absolute correspondence of the pitch elicited by acoustic F0 and electric pulse rate for conditions that afford good electric rate-pitch salience.

The relevance of the outcomes of this study to sound coding strategies is twofold. Firstly, for existing clinical strategies the elicited pitch for low F0 signals is expected to be substantially higher on average and its growth with increasing F0 shallower than in normal hearing. The relationship between F0 and pitch is also likely to be inconsistent across different sound sources and acoustic environments. Secondly, for strategies that explicitly code F0 information using deep SAM or some other modulation shape, such as EDM, the relationship between F0 and pitch is expected to be more normal for low F0 signals. Although an upper limit to electrical rate discrimination is expected regardless of the modulation shape, that limit is expected to be lower for shallow (or broad) modulators. In a practical implementation, the reduction in loudness that arises from application of deep F0 modulation in the stimulus envelope may need to be compensated for, particularly for low F0s and for EDM.
4 Development of a Rate-Pitch Strategy (eTone)

Based on findings of previous research and the experiments conducted in chapter 3, a sound coding strategy for users of cochlear implants was developed to improve coding of fundamental frequency (F0) in the temporal envelopes of the electrical stimulus signals. The strategy, known as Enhanced-Envelope-Encoded Tone (eTone) is based on the Advanced Combinational Encoder (ACE) strategy (§2.1) and includes additional processing that explicitly applies F0 modulation to channel envelope signals that contain harmonics of prominent complex tones. Channels that contain only inharmonic signals retain envelopes normally produced by ACE. The strategy incorporates an F0 estimator to determine the frequency of modulation and a harmonic probability estimator to control the amount of modulation enhancement applied to each channel. The F0 estimator was designed to provide an accurate estimate of F0 with minimal processing lag and robustness to the effects of competing noise. Error rates for the F0 estimator and accuracy of the harmonic probability estimator were compared with previous approaches and outcomes demonstrated that the strategy operates effectively across a range of signals and conditions that are relevant to cochlear implant users.

4.1 Introduction

Perception of voice pitch is important in everyday life as it provides information about the identity, gender and emotional state of a speaker, distinguishes questions from statements, and is useful in separating concurrent sound sources (e.g., Assmann and Summerfield, 1990). For tonal languages, such as Mandarin and Cantonese, a change in voice pitch is used to convey lexical meaning (§2.5), and in music pitch determines the musical intervals that form melodies (§2.4). Over the past thirty years significant advances in cochlear implant technology have resulted in incremental improvements in hearing abilities for people with severe to profound deafness, and many implant users show good abilities related to speech recognition. However, voice-pitch and music perception to date remains far from satisfactory (§2.4.1; §2.5).

In normal hearing, pitch is primarily encoded via fine spectro-temporal information in apical auditory filters in which harmonics of the fundamental frequency are resolved (e.g., Dai, 2000, also see §2.2.1). Pitch can also be encoded via temporal envelope information in wider auditory filters where unresolved harmonics beat with one another (Houtsma and Smurzynski, 1990). In CI hearing, the broad neural activation that results
from electrical stimulation precludes coding of fine spectral information necessary to resolve harmonics (§2.3.1). In addition, rate-pitch limitations in CIs (§2.3.2) means that fine temporal information beyond approximately 300 Hz can not be conveyed (e.g., Zeng, 2002). However, this rate limitation does not preclude coding of temporal envelope cues to pitch for F0s up to approximately 300 Hz where rate or modulation rate of electrical pulse trains can be used to convey a percept of pitch (§2.3.2; §2.3.3).

Numerous CI strategies that code F0 pitch information in addition to speech have been developed over the past decades (§2.6.2). The earliest approaches coded F0 via rate of stimulation, such as the “F0/F2”, “F0/F1/F2”, and “Multipeak” strategies. Speech recognition with those strategies was poor compared to present day systems, particularly in noisy situations for which estimation of speech features and F0 was difficult. Later devices coded F0 information via amplitude modulation in the stimulus envelope that was related to F0, e.g., the CIS, SMSP, SPEAK, and ACE strategies (Wilson et al., 1991; McKay et al., 1991; Skinner et al., 1994; and Vandali et al., 2000 respectively; §2.1). Whilst good speech recognition has been demonstrated using those strategies, pitch perception has been shown to be poor (Geurts and Wouters, 2001; McDermott, 2004; Vandali et al., 2005). These results have been attributed mainly to the shallow and uncontrolled depth of F0 modulation provided in the stimulus envelope of these strategies.

More recently, researchers have examined algorithms capable of enhancing temporal cues to F0 by either expanding the depth of amplitude modulation in channel signals or by explicitly modulating channel signals at an F0 rate (§2.6.3). For instance, Geurts and Wouters (2001) devised a CIS-based strategy in which the depth of envelope modulation in each channel was increased (§2.6.3.1). Zero-phase band-pass filters were employed to minimise distortion of the temporal envelopes across channels (Oppenheim and Schafer, 1989). However, no significant benefit as compared to CIS was obtained in F0 discrimination tests using synthetic vowel stimuli. The present authors have examined a number of similar strategies (§2.6.3.3) designed to expand the depth of envelope modulations in channel signals. For instance, the Multi-channel Envelope Modulation (MEM) strategy (Vandali et al., 2005) modulated low-pass filtered channel signals of ACE by the broadband envelope of the input signal which had been processed to expand the depth of any F0 modulation in the range of 80 to 300 Hz. This processing provided deeper modulation cues to F0 which were phase-aligned across all activated electrodes. Significant improvements in pitch ranking of sung-vowels separated by half
an octave were obtained compared to ACE. In addition, average speech recognition scores in quiet and noise using MEM were no worse than with ACE. However, a number of problems associated with these types of strategies have been identified. Firstly, the frequency of temporal envelope information in band-pass filtered or broadband signals is not always equal to F0 because the temporal waveform can contain multiple peaks per F0 period. Secondly, modulations in the temporal envelope are easily corrupted by noise. Current research in this area is now focused on improved methods to explicitly estimate F0 to control electrical pulse rate or modulation rate.

In a series of studies, Green et al., (2004, 2005) investigated the performance of a non-real-time CI strategy that modulated the band-pass filtered, slow-rate (i.e., 32 Hz low-pass filtered) envelope signals of CIS by a sawtooth waveform (§2.6.3.2). The frequency of the sawtooth waveform was matched to the estimated F0 of the input signal. Results compared to CIS demonstrated modest but improved identification of pitch direction in processed synthetic diphthongal glides and in the ability to use intonation to identify sentences as questions or statements. However, a reduction in vowel recognition was also observed. The authors indicated that whilst the processing provided improvements to temporal pitch perception, it adversely affected transmission of spectral information. In addition, they noted the need for a real-time F0 estimator that was robust to the effects of noise.

In a study by Laneau et al., (2006), a real-time F0 modulation strategy was compared to ACE in a group of six CI users (§2.6.3.4). The strategy employed an autocorrelation of the power spectrum to estimate F0, and modulated the band-pass filtered, slow-rate envelope signals of an ACE like strategy by a sinusoidal waveform of frequency equal to the estimated F0. Improved F0 discrimination for synthesised musical notes played by five different instruments was observed for F0s up to 250 Hz. In addition, improved melody recognition of familiar Flemish songs (with rhythm cues removed) was shown. The authors however noted that an assessment of any adverse strategy effects on speech intelligibility was needed and that a more accurate, robust, and efficient F0 estimation technique was required before the strategy could be successfully applied in a clinical system.

The work described in this paper expands on these previous studies and ideas. A sound-coding strategy is presented that is based on ACE and incorporates additional features designed to enhance coding of F0 information in the stimulus envelope. The main objective in this strategy is the provision of F0 modulation in the stimulus
envelope, synchronised across all frequency channels, for channels that contain one or more harmonic(s) of F0. For channels that are not related to F0, the stimulus envelope at the output of the narrow-band filter associated with that band is preserved as per the normal ACE strategy. Because real signals often contain a time-varying combination of harmonic and inharmonic components distributed across the frequency spectrum, the strategy is required to dynamically determine relative amounts of F0-modulated and non-F0-modulated signals in each channel from the degree to which the channel signal is harmonically related to F0. The performance of such a strategy depends on how well F0 can be estimated and thus care was taken to develop a real-time F0 estimator that was accurate, had minimal lag, and was robust to the effects of competing noise. Finally, the shape of the F0 modulation applied to the stimulus envelope was chosen to provide deep F0 modulation and a low-duty cycle so that F0 was well represented in neural response timing. A detailed description of a real-time strategy that meets these requirements follows along with an analysis of its performance for a range of speech and sung vowel stimuli of varying F0 and signal-to-noise ratios (SNRs).

4.2 Methods

4.2.1 The eTone strategy

The Enhanced-Envelope-Encoded Tone (eTone) strategy (Vandali and van Hoesel, 2009; 2010) is based on the Advanced Combinational Encoder (ACE) strategy (Vandali et al., 2000), which is one of the most widely used clinical strategies presently available to CI users. To compare performance for the two strategies, both were implemented in the SPEAR3 sound processor which is a portable device suitable for CI research. Referring to Fig. 4.1, both strategies divide the incoming sound signal (sampled at 16 kHz) into channel signals (typically $N_{ch} = 20$) using a bank of band-pass filters (BPFs) and envelope estimators which span a frequency range of approximately 125 to 8000 Hz. The bank of BPFs is implemented using a 128-point Fast Fourier Transform (FFT) in which complex addition (equivalent to the Hilbert Transform) of FFT bin vectors is used to construct the channel envelope signals. Typically for ACE, up to eight to ten of the BPF channel signals $E$ that have the highest level are then selected as “maxima” and are used to derive the stimulus intensities for subsequent electrical activation of intra-

---

7 The SPEAR3 processor is a research system developed by the Hearing CRC. Information about this device can be obtained from HearWorks, 550 Swanston st, Carlton, 3053, Australia, www.hearingcrc.org/commercialisation/hearworks.
cochlear electrode sites that correspond in a tonotopic manner to the channel frequencies. The stimulus intensities of the activated electrodes are derived from a loudness-growth function that maps channel intensity to stimulation levels within the CI user’s range of threshold and maximum comfortable loudness levels of stimulation. The stimuli are encoded into radio-frequency (RF) signals and transmitted to the implanted receiver/stimulator. The electrodes are activated sequentially, at a fixed stimulation rate that is equal to the rate at which the channel signals are analysed. A rate of 1455 Hz per channel was employed in the present implementation.

Figure 4.1. Block diagram of the experimental strategy showing the signal path from audio input through to the cochlear implant. A switch on the SPEAR3 processor is used to select ACE or eTone processing. Processing blocks added by eTone are highlighted in yellow.

The eTone strategy provides additional processing to ACE that enhances coding of harmonic/periodic sounds. Essentially, channel envelope signals are smoothed by an envelope tracker to remove any F0 modulation. For an input signal that consists of a dominant harmonic tone, the smoothed channel envelope signals $E_S$ are modulated at a rate determined by a real-time F0 estimator. The modulated channel signals and the non-modulated (ACE) channel signals $E$ are then combined based on the degree to which each channel signal is related to F0 (as derived by the harmonic probability estimator). For channels which are strongly related to F0, a high ratio of the F0-modulated-to-non-modulated signals is used whereas for those that are weakly related, a low ratio is used. These combined signals $E_M$ are then passed to the maxima selector and processing continues as per the standard ACE strategy. The following sections provide a detailed description of the processing stages employed in eTone.
4.2.1.1 \textit{F0 estimator}

A real-time F0 estimator is used to estimate the fundamental frequency $F_{0_{\text{EST}}}$ of the most dominant harmonic signal in the input. The F0 estimator is based on the “harmonic sieve” developed by Duifhuis et al., (1982). A modified version of this device was described in Zakis et al., (2007). The implementation employed in the present study adds considerable sophistication distributed over six processing stages and is illustrated in Fig. 4.2.

Figure 4.2. Block diagram of F0 estimator. The example input is illustrative of the vowel /a/ with an F0 of 210 Hz. In stage 1, the power and frequency of components in the signal are determined. In stage 2, the signal is passed through a bank of harmonic sieves (only sieves 4 to 16 are shown which span one octave of F0s from 103.8 to 207.7 Hz). The power passed (matched) by each sieve is determined and the largest matched powers are used to select candidate F0 sieves 4, 7, 9, and 16. For each candidate sieve, the F0 of the signal contained is estimated. The estimated F0 ($E_{F0}$) is used in the third stage to determine band centre-frequencies for a narrower band sieve. The finer frequency resolution of these bands reduces the matched power passed by sieves 7 and 9. In stage 4, an estimate of the noise (or inharmonic) power is subtracted from the matched power passed by the narrow-band sieves, which also reduces the matched power in sieves 7 and 9. In stage 5, a weighting function is applied to the candidate sieve matched powers which emphasises higher F0 frequencies to reduce octave errors, thus resulting in greatest matched power for sieve 16. In the final stage, the best F0 estimate is selected from those obtained over a number of consecutive time frames.
The first stage estimates the power and frequency of components in the input signal up to a frequency of around 2 kHz which covers much of the harmonic frequency range useful for perception of pitch (Dai, 2000). In the second stage candidate F0 values are determined by comparing the amount of power present (or matched to F0) in a series of harmonic sieves. Each sieve consists of a bank of filters with harmonically related centre frequencies located at integer multiples of its F0. Filters are rectangular with a bandwidth of one semitone. Sieve F0 frequencies are separated by one semitone from approximately 80 to 400 Hz which covers the range of rates that can be perceived by typical CI users (McDermott, 2004). Sieves with the largest power output are selected. For each case a candidate F0 ($E_{F0}$) is determined and is further evaluated in stages three-six. These later stages are used to minimise F0 errors which mainly manifest as octave errors. In noisy conditions, the coarse frequency resolution of harmonic bands employed in stage two can lead to erroneous addition of noise (or inharmonic) power to the matched power estimate, particularly in high-order bands of low frequency candidate F0s. Thus in the third stage, the signal is subjected to a second bank of harmonic sieves that have much narrower bands centred at integer multiples of the candidate F0 estimates. In the forth stage, which is optional, an estimate of the noise (inharmonic) power is subtracted from the matched power passed by the narrow-band sieves. Stage five is used to minimise F0 octave errors which arise because the matched power in sieves that share the same harmonic bands (i.e., that are octave related) will be equal or similar. A weighting function is applied to the matched powers passed by the narrow-band sieves so that higher F0 frequency sieves are given greater weighting than lower ones. The candidate F0 sieve with the highest weighted power at the output of the fifth stage is then selected as the F0 estimate for the current time frame. The final stage is used to further reduce spurious F0 estimation errors in noisy conditions by selecting the best F0 estimate from those obtained over a number of consecutive time frames. A detailed description of each stage follows.

**Stage 1:** A phase-vocoder FFT filterbank (Moore, 1990) is utilised to provide estimates of the frequencies and powers of sinusoidal components present in the input. The input from the SPEAR3 behind-the-ear (BTE) microphone is initially low-pass filtered (LPF) using a 2.2 kHz, 4th order, Butterworth filter, and down-sampled to 8 kHz. The most recent 32 ms (256 samples) of the signal are windowed using a Hanning window (which provides a -3 dB bandwidth of 31.25 Hz). A “stacking and adding” technique (Crochiere and Rabiner, 1983) is used to reduce the number of samples to 128
Chapter 4: Development of a Rate-Pitch Strategy (eTone)

via decimation in the frequency domain by two\(^8\). These samples are circularly rotated in time by \(-k \times t\) samples, where \(t\) = analysis frame number, and \(k\) = analysis frame rate = 11 samples (giving an analysis frequency = \(8000/11 = 727\) Hz). The rotation is performed so as to maintain zero relative phase-shift between successive analysis frames in the phase-vocoder. A 128-point FFT is performed to obtain an estimate of the complex frequency spectrum, where FFT bin numbers \((b)\) 1 to 32, correspond to frequency bands spaced by 62.5 Hz from 62.5 to 2000 Hz. Bin power values are derived from the sum of squared real and imaginary FFT values. Bin frequency values are estimated from the phase difference between successive FFT frames (c.f., Moore, 1990).

Average bin powers \(p_b\) and a power-weighted average of bin frequencies \(f_b\) are determined from four successive phase-vocoder frames (at \(727/4 = 182\) Hz) to smooth any spurious values. The allowable upper value for average bin frequencies is \(MaxF = (32 + 0.5)\) FFT bins \(\times 62.5\) Hz bin-width = 2.031 kHz. Average bin powers with frequencies beyond this limit are set to zero. The 0-2 kHz frequency response of the system is flat for signals fed directly into the SPEAR3 processor. However for signals presented in the sound field, approximately +4 dB/octave of emphasis is introduced by the BTE microphone over the 2 kHz frequency range. Thus, -4 dB/octave de-emphasis is applied to average bin powers for signals presented in the sound field.

Stage 2: F0 estimates are determined using a range of F0 sieves starting from 82.4 Hz (E2 in a western musical scale) and increasing in steps of one semitone (5.94\%) up to some upper limit. In the present implementation, the upper F0 is 391.9 Hz (G4). The amount of power in the signal that is harmonically related to each sieve’s F0 is the matched power \(MP\). It is calculated for each sieve according to Eq. (4.1) by summing all average bin powers \(p_b\) for which bin frequencies \(f_b\) fall within a harmonic series of rectangular bands (i.e., ideal BPFs) centred at multiples of the sieve’s F0 frequency. The centre frequencies for each sieve \(F_{r}[c]\) are described by Eq. (4.2), for \(c = 1\) to 28, where \(c\) is the sieve number and \(C_{F0}\) is the sieve F0 frequency = \(82.4 \times 2^{c/12}\) Hz. Each band spans ±0.5 semitones around all integer multiples \((i)\) of \(C_{F0}\) up to a maximum harmonic

\(^8\) Frequency decimation halves FFT size by omitting odd numbered bins whilst maintaining bin bandwidth. This provides higher frequency resolution for resolving harmonic frequencies at the expense of introducing ripple in the composite frequency response. A method of restoring the composite frequency response is employed after the phase-vocoder which scales bin powers by the inverse magnitude response of each FFT bin at the bin frequency derived from the phase-vocoder. Only bin powers with bin frequencies within ±50 Hz of the bin centre frequency are adjusted.
Chapter 4: Development of a Rate-Pitch Strategy (eTone)

frequency of MaxF. A small amount of overlap between harmonic bands of adjacent F0 sieves is provided by using $F_{offset} = 2$ Hz.

$$M_p[c] = \sum_b p_b$$  \hspace{1cm} (4.1)

$$F_i[c] = \{f : f \geq 2^{-0.5/12} i C_{F0} - F_{offset} \text{ and } f \leq 2^{+0.5/12} i C_{F0} + F_{offset}\}$$  \hspace{1cm} (4.2)

F0 sieves with the largest matched power (i.e., those within -3 dB, or half, of the largest matched power) are selected as candidate F0 sieves. For each selected sieve, an estimate of the F0 frequency $E_{F0}$ is derived in Eq. (4.3). For each harmonic band in the sieve, an associated F0 value is determined by dividing the bin frequency by the harmonic number $i$ and an overall F0 estimate $E_{F0}$ is determined from a power weighted average of band F0 estimates.

$$E_{F0}[c] = \sum_b p_b \frac{f_b}{i} \left/ \sum_b p_b \right.$$  \hspace{1cm} (4.3)

Stage 3: For the candidate F0 sieves with the largest matched power selected in stage two, a second harmonic sieve process is employed which uses narrower harmonic bands centred at integer multiples of each candidate F0. The harmonic bands are Gaussian in shape, as described by Eq. (4.4), with mean centre frequencies positioned at harmonic multiples of $E_{F0}$. The bandwidth of Gaussian bands is controlled by the standard deviation $K_G$. For $K_G = 0.02 \times C_{F0}$, $G = 0.5$ (i.e., -3 dB power bandwidth) when the bin frequency divided by the harmonic number (i.e., $f_b/h$) is approximately ±2.4% (or ±0.4 semitones) away from $E_{F0}$. In this case, the harmonic number $h$, described by Eq. (4.5), is the integer multiple of the estimated F0 frequency $E_{F0}$ that is closest to the bin frequency $f_b$.

$$G(f_b, E_{F0}[c]) = \exp \left( \frac{- (f_b/h - E_{F0}[c])^2}{2K_G^2} \right)$$  \hspace{1cm} (4.4)
Chapter 4: Development of a Rate-Pitch Strategy (eTone)

\[ h = \left\lfloor f_b / E_{F0}[c] + 0.5 \right\rfloor \]  

(4.5)

For low-frequency candidate F0 sieves, high order harmonic bands overlap significantly if \( K_G \) is too high. For this reason, and to limit noise contributions in wide bands at high frequencies, maximum bandwidth \( BW_{Max} \) is limited to 60 Hz by limiting \( K_G \) to \( K_{GMax} \) according to Eq. (4.6).

\[ K_{GMax}(h) = \frac{BW_{Max}}{2h\sqrt{-2\ln(0.5)}} \]  

(4.6)

The amount of matched power passed by the narrow-band sieves is determined as per Eq. (4.7) by summing the bin powers weighted by the Gaussian function over a set of frequencies \( F0_r[c] \) that spans ±2 semitones around each integer multiple of \( E_{F0}[c] \).

\[ \forall b; f_b \in F0_r[c] \]  

\[ M_p[c] = \sum_b p_b \cdot G(f_b, E_{F0}[c]) \]  

(4.7)

Stage 4: For monophonic signals and ideal/quiet conditions, the matched power consists entirely of harmonic signal power. However in noisy conditions, the matched power comprises both the harmonic signal power and a portion of the noise (or inharmonic) power. A better estimate of the harmonic signal power can optionally\(^9\) be derived by subtracting an estimate of the noise power within the sieve from the matched power. If it is assumed that the input signal consists of a monophonic complex harmonic signal and noise is distributed uniformly across the 2 kHz range, the contribution of the noise power \( N_P \) to the total harmonic matched power \( M_P \) increases as the summed bandwidth of the sieve filters \( S_{BW} \) increase relative to the total bandwidth \( T_{BW} = MaxF - 60 \) Hz. It can be shown that:

\[ N_P = (T_p - S_p) \times K_{BW} \times S_{BW}/T_{BW} \]  

(4.8)

\(^9\) The stage four option was employed in the reported implementation of the F0 estimator. However, if this stage is omitted, all references to “signal power” in subsequent stages of the F0 estimator should be replaced by “matched power”.

118
where $S_P$ represents the improved estimate of harmonic signal power, $T_P$ is the total power over $T_{BW}$, and $K_{BW}$ is used to compensate for the fact that signal and noise are typically not uniformly distributed across the frequency range. In examining variations of $K_{BW}$ it was found that a value of 0.5 provides a good compromise between noise power estimates for different signals, noise types, and SNRs. The sieve bandwidth $S_{BW}$ is calculated as the sum of the -3 dB power bandwidths of each Gaussian filter in the sieve:

$$S_{BW}[c] = \sum_{h} 2hK_{G}\sqrt{-2\ln(0.5)} \quad (4.9)$$

Given that $M_P = S_P + N_P$, transposition of Eq. (4.8) yields the improved estimate of the harmonic signal power as:

$$S_P[c] = \frac{M_P[c] - T_PK_{BW}\left(\frac{S_{BW}[c]}{T_{BW}}\right)}{1 - K_{BW}\left(\frac{S_{BW}[c]}{T_{BW}}\right)} \quad (4.10)$$

**Stage 5**: The next stage of processing involves minimisation of octave errors, which are problematic in many F0 estimators. For harmonic sieve based estimators, octave errors arise because all harmonics of F0 are also harmonics of integer submultiples of F0 (i.e., lower octaves of F0). Therefore at the output of stages two, three, and four, equal matched power would for example be obtained in quiet conditions for candidate F0s corresponding to the actual signal F0, and half that frequency (as is the case for the example in Fig 4.2). This problem can be counteracted by applying a small amount of positive weighting to matched powers of higher candidate F0s. However, too much positive weighting can introduce the opposite error, in which a higher octave F0 is incorrectly estimated when the amount of energy of odd numbered harmonics of F0 is low compared to that of even numbered harmonics. Careful choice of the weighting function is required to minimise both errors. The choice of weighting function is further dependant on SNR. As noise is introduced, the matched power for lower candidate F0s increases more than for higher candidate F0s (because they sum power from more
harmonic bands) and thus greater positive weighting is needed for higher candidate F0s to counteract sub-octave F0 errors.

In the current implementation, the weighting function is used to compensate for difference in the combined bandwidth of all harmonic bands summed (i.e., the sieve bandwidth $S_{BW}$) for each sieve. It is inversely proportional to the sieve bandwidth raised to the power $K_W$, as given by Eq. (4.11), where the constant $K_W$ is used to adjust the degree of positive weighting.

$$W[c] = S_{BW}[c]^{-K_W}$$  \hspace{1cm} (4.11)

The weighted harmonic signal power is calculated as $WS_P[c] = S_P[c] \times W[c]$. The candidate F0 with highest weighted harmonic signal power\footnote{Only candidate F0 sieves for which at least two harmonics are summed are considered when determining the largest weighted signal power.} is determined and its estimated F0 ($E_{F0}[c]$) is used as the F0 estimate for the current frame of the F0 estimator. Through experimentation using a range of speech signals with various F0s in quiet conditions, best F0 estimation accuracy was obtained for $K_W = 0.02$ to $0.1$. However when noise was added to the signal, higher values of $K_W = 0.20$ to 0.34 were required to compensate for the increase in noise power summed by low candidate F0 sieves. Thus, a user controllable switch on the SPEAR3 processor is used to select between “quiet” or “noise” processing modes for which $K_W$ is set to 0.08 or 0.3 respectively. Typically, for high-to-moderate SNRs (i.e., greater than approximately +6 dB) the quiet processing mode is selected, for lower SNRs the noise processing mode is selected. An algorithm for adaptive adjustment of $K_W$ was also examined. It estimates the harmonic signal-to-total power ratio $STR$ for the largest signal power derived in stage four using $STR = S_P / T_P$. The $STR$ value can range from 1.0 corresponding to a high SNR, through to approximately 0.5 or lower for SNRs of 0 dB and lower. The adaptive algorithm linearly adjusts $K_W$ between values of 0.02 to 0.34 for $STR$ values ranging from 1.0 to 0.55. For $STR$ values less than 0.55, $K_W$ is limited to 0.34.

Stage 6: In noisy conditions, F0 estimation errors are further reduced by selecting the best F0 estimate from a number of consecutive estimator frames using a process somewhat resembling “listening in the gaps”. For each frame, the largest weighted signal power is normalised by the total power to provide an estimate of the weighted
signal power-to-total power ratio \( WSTR = WSP / TP \). Through experimentation using different signals, noise types, and SNRs, it was found that the highest \( WSTR \) over several frames provides a more robust indicator of F0 than does the highest weighted signal power, and it is therefore used to select the frame that determines the final F0 estimate \((F0E_{ST})\). The number of consecutive frames sufficient for accurate F0 estimation at high SNRs (and in quiet) was around 4, whereas for low SNRs best results were obtained by using up to 12 frames. The switch on the SPEAR3 processor used to select between “quiet” and “noise” processing modes therefore also sets frame buffer sizes to 4 or 12 respectively (even when using the adaptive \( K_W \) algorithm). An estimate of the “unweighted” signal-to-total power ratio \( STR = SP / TP \) (i.e., without application of the weighting function in stage five) for the selected frame is also determined. This value, referred to as the broadband \( STR \), is used in later stages of the strategy as a measure of the F0 estimation strength, or the probability that the signal (within the 0-2 kHz range) is harmonic. Finally, FFT bin power \( p_b \) and frequency \( f_b \) values for the best F0-estimate frame are passed to the harmonic probability estimator stage as \( P_{F0}[b] \) and \( F_{F0}[b] \) values respectively.

The update rate for this implementation of the F0 estimator is 182 Hz (or 5.5 ms) and the total processing delay is 18 ms for the quiet processing mode or 40 ms for the noise processing mode. The number of DSP program cycles/second required for its operation is of the order of 10 MIPS, which is approximately 80% more operations/second than that required to implement the ACE coding strategy. As implemented, the algorithm requires approximately 5 K words of memory space and 3 K words of program space. Whilst no direct comparisons to other F0 estimators were carried out, we estimate computational complexity to be up to twice as high as techniques based on auto-correlation. The increased complexity is likely to impact on power consumption and battery life.

4.2.1.2 Wide-bandwidth band-pass filterbank

A second filterbank stage, in parallel to the normal (ACE) filterbank, is included to determine the wide-bandwidth channel envelope signals \( E_{WB} \) for channel frequencies above \( MaxF \) (2 kHz). Its output is used by the harmonic probability estimator to determine the probability that the channel signal is related to the estimated F0. It uses the same FFT outputs, and has identical BPF centre frequencies (CFs) as those used in the normal filterbank. However, each BPF has a minimum bandwidth sufficient to pass at
least two F0 harmonics of the highest F0 to be analysed by the system (i.e., at least approximately 800 Hz wide for a maximum F0 of 400 Hz). Thus for harmonic tones, these channels will carry amplitude modulation in their envelope related to F0.

### 4.2.1.3 Harmonic probability estimator

For each channel, the harmonic probability estimator determines the probability that the channel envelope signal is related to the estimated F0. Two methods are used to achieve this. Firstly, for BPF channels within the frequency range analysed by the F0 estimator (i.e., up to 2 kHz which typically correspond to channels \( n = 1 \) to 11 for a 20 channel filterbank), the probability that the channel signal contains sinusoidal components that are an integer multiple of the estimated F0 is determined. For each channel the harmonic signal power within the corresponding BPF is estimated by weighting all bin powers \( P_{F0}[b] \), within the frequency range \( F_{ch} \) corresponding to the -60 dB power bandwidth of the BPF, according to how closely their frequencies match that of harmonics of the estimated F0. The weighting factor is calculated using the Gaussian function described in Eq. (4.4) where in this case \( K_G = F0_{EST} \) and \( BW_{Max} = 25 \text{ Hz}^{11} \). This results in 25 Hz wide Gaussian filters for all harmonics and F0s. The weighted bin powers are scaled by the power response of the BPF channel \( P_{ch} \) for each bin frequency \( FF0[b] \) and summed to provide an estimate of the F0 related signal power within the channel. That power is divided by the total power in the channel to determine the channels’ harmonic-to-total power ratio \( STR_{ch} \) as per Eq. (4.12).

\[
STR_{ch}[n] = \frac{\sum_{b:F_{ch}[b] \in F_{ch}} P_{F0}[b] G(F_{F0}[b], F0_{EST}) P_{ch}(F_{F0}[b])}{\sum_{b:F_{ch}[b] \in F_{ch}} P_{F0}[b] P_{ch}(F_{F0}[b])} \tag{4.12}
\]

The channel \( STR_{ch} \) is normalised (i.e., multiplied) by the broadband \( STR \) (which reflects the probability that the 0-2 kHz signal is harmonically related to the estimated F0). The normalised channel \( STR_{ch} \) ranges from approximately 1 when the channel signal contains frequency components related to the estimated F0 and the broadband

---

11 Alternative values of \( K_G \) and \( BW_{Max} \) can be used to adjust the tolerance (filter bandwidths) of the harmonic probability estimator. For instance, higher values of \( BW_{Max} \) will increase the tolerance. Setting \( K_G = 0.15 \times F0_{EST} \) will decrease the tolerance for low-order harmonics.
Chapter 4: Development of a Rate-Pitch Strategy (eTone)

STR is high, down to approximately 0 when the channel signal is not related to the estimated F0 or STR is low. A channel harmonic probability value is then determined by transforming the normalised \( STR_{ch} \) value using a sigmoidal function as per Eq. (4.13), where \( a = 0.5 \) sets the inflection point, and \( b = 0.5 \) sets the 5%-95% width of the sigmoid function. This function approaches 1 for scaled \( STR_{ch} \) values of 0.75 or higher, 0 for scaled \( STR_{ch} \) values of 0.25 or lower, and values between 1 and 0 for intermediate \( STR_{ch} \) values.

\[
HP_{ch}[n] = \left(1 + \exp\left(-6 \times \frac{STR_{ch}[n] \times STR - a}{b}\right)\right)^{-1} \tag{4.13}
\]

Because the channel harmonic probability determines the modulation depth applied to the envelope by the channel modulator stage, it is reduced when the channel envelope level is within the lower 16 dB of the channel input dynamic range (DR, which is typically set to 40 dB) to avoid excessive reductions in the loudness at low channel intensities. The harmonic probability is scaled by a factor varying from 1.0 to 0.25 over this 16 dB range. As a final stage it is filtered using a 30 Hz, 1st order, LPF to smooth spurious fluctuations.

The second method of estimating the channel harmonic probability is employed for channels above the frequency range analysed by the F0 Estimator (i.e., > 2 kHz for \( n = 12 \) to \( N_{ch} \) typically). In each channel, the wide-bandwidth envelope signal \( E_{WB} \) is high-pass filtered (100 Hz, 1st order) and the autocorrelation is calculated for a time shift of \( \tau = 0 \), as well as for \( \tau = 1/F0_{Est} \) over a 27.5 ms window. The ratio of the F0-period time shifted auto-correlation value to the zero time shifted value describes the channel harmonic probability as per Eq. (4.14), where \( HPF \) is the high pass filter function and \( ACF(f, l) \) is the auto-correlation function of \( f \) for a lag \( l \).

\[
STR_{ch}[n] = \frac{ACF(HPF(E_{WB}), 1/F0_{Est})}{ACF(HPF(E_{WB}), 0)} \tag{4.14}
\]

As in the first method described above, the channel \( STR_{ch} \) value is normalised by the broadband \( STR \) and transformed using a sigmoid function (as per Eq. 4.13 with \( a = 0.35 \) and \( b = 0.5 \)), to give the channel harmonic probability value. It is then attenuated
when the slow-varying channel envelope level is within the lower 16 dB of the channel DR and finally low-pass filtered.

### 4.2.1.4 Envelope tracker

An envelope tracker is employed to determine the slow-varying channel envelope signals $E_S$. These signals are used by the channel modulator stage for generation of the F0 modulated channel signals. The envelope tracker is designed to follow the peak level of the signal and remove modulations above approximately 73 Hz. It achieves this by using an instantaneous attack time to follow peaks in the channel envelope signal $E$. Peak signal levels are held (using a very slow, 0.02 Hz, 1st order, LPF) for a period of 20 samples (or 13.75 ms which is slightly longer than the period of the lowest F0 signal that can be estimated by the strategy) unless a larger signal level is encountered, in which case the hold time is reset. After completion of the hold-time, a fast release time (300 Hz, LPF) is applied until the signal again exceeds the slow-varying envelope level. A processing delay of 13.75 ms is introduced in the signal path by the envelope tracker. The same delay is also introduced in the channel envelope $E$ signal path so that both paths are aligned in time.

### 4.2.1.5 Channel modulator

The channel modulator stage is used to apply F0 modulation to the slow-varying channel-envelope signals and to combine (or mix) these signals with the non-modulated (ACE) channel envelope signals. A modulation function $M$, with modulation frequency equal to the estimated F0, is used to modulate the slow-varying channel envelope signals $E_S$. The modulated signals are scaled by the channel harmonic probabilities $HP_{ch}$. A modulation gain constant $K_{HP}$ is included to optionally increase the level of the modulated signals so that the loudness elicited by the signal can be approximately matched to that of unmodulated signals. The unmodulated channel envelope signals $E$ are scaled by the channel non-harmonic probabilities (i.e., $1 - HP_{ch}$) and mixed with the F0 modulated channel signals as per Eq. (4.15), for channels $n = 1$ to $N_{ch}$. The total modified channel envelope signals $E_M$ are then passed to the maxima selection stage where processing continues as per the standard ACE strategy.

$$E_M[n] = (HP_{ch}[n] + K_{HP}) E_S[n] M(F0_{EST}) + (1 - HP_{ch}[n]) E[n]$$  \hspace{1cm} (4.15)
The modulation function $M$ is designed to have a low duty cycle so that most of the stimulation current in each F0 interval is imparted by a single electrical pulse. This is done because previous studies have demonstrated that the pitch elicited by electrical pulse trains is governed by a function of the longest first-order intervals between pulses rather than the period of modulation (e.g., Carylon et al., 2002). In addition the results reported in chapter 3 for sharp-onset exponential-decay (EDM) modulated stimuli, demonstrated close pitch matches between EDM modulated stimuli and unmodulated pulse trains and when modulation rate was equal to the pulse rate. For the present study, the modulation function is similar to EDM. It consists of a narrow pulse with an instantaneous attack and exponential decay. Control is provided to adjust the depth of the modulation function and its exponential decay rate. By default, the modulation depth $MD$ (defined as peak/trough stimulation level in clinical current units) is adjusted to 50% of a subject’s electrical DR which translates to an acoustic-equivalent depth of 20 dB given a 40 dB DR in each channel. The default exponential decay function falls to 10% of its peak value within the first 25% of the modulation period. The modulation function (stored as a single cycle in 128 point buffer) is sampled at intervals of $(F0_{EST} \times \frac{128}{\text{stimulation rate}})$ samples. However, because the stimulation rate can be a non-integer multiple of F0, amplitude beating in the sampled output can arise. To avoid this, at the beginning of every F0 modulation cycle, sampling of the modulation function is reset to the first sample of the modulation function. The start of each F0 cycle is determined by maintaining an accurate ongoing record of the desired F0 modulation phase.

4.2.1.6 Maxima selector

A modification to the maxima selection process was included to improve selection of channels containing harmonics of F0, particularly in noise conditions. Prior to maxima selection, additional gain proportional to the channel harmonic probabilities is applied to the modulated component of channel signals (in a similar way to that provided by the modulation gain constant $K_{HP}$). After channels containing maxima have been selected, the channel levels are restored to their unadjusted values. For the quiet processing mode, up to +6 dB of gain is applied whereas for the noise mode up to +20 dB of gain is applied. This processing increases the likelihood that F0 modulated channels are selected in preference to noisy/unmodulated channels, ensuring that the coded envelope of the F0 modulation remains more contiguous in selected channels. However, for poor SNRs it is
possible that the envelope of the F0 modulation in some channels will be disrupted by selection of higher level noise in un-modulated channels.

4.2.2 F0 estimator analysis

Measurements of F0 estimator error rates were performed using sung-vowels and sentences sung/spoken by male and female speakers covering a range of F0s appropriate to temporal pitch coding in CIs, (i.e., up to at least 300 Hz). For the sung-vowel stimuli (Sucher and McDermott, 2007) the vowels /a/, as in “hard”, and /i/ as in “heed” were analysed. The vowel /a/ has the highest first formant frequency (F1) and a low second formant frequency (F2), whereas /i/ has the highest F2 and a low F1. For the male singer, four repetitions of each vowel sung at 22 F0s ascending by one semitone from G2 (98 Hz) to E4 (329.6 Hz) were analysed. For the female singer, four repetitions of each vowel sung at eight F0s ascending from C4 (261.6 Hz) to G4 (392 Hz) were analysed. High resolution spectral analysis of each vowel was carried out to determine actual average F0 values, and differences between measured F0 values and those of the standard Western musical scale were usually less than 1.0% and never exceeded 1.39% (Zakis et al., 2007). The duration of each vowel token was 560 ms which included a 30 ms linearly ramped onset and offset. All tokens were equalised to a fixed average RMS level. The vowels were presented in quiet, in white noise, and in eight-talker babble at SNRs between +12 dB and -12dB in 4 dB steps (measured using flat frequency weighting across the 0-8 kHz frequency range). Error rates for the F0 estimator were calculated from the number of erroneous F0 estimates divided by the total number of estimates summed over the entire duration of the vowels. F0 estimates were considered correct if they fell within a criterion of ±0.5 semitones (or ±2.9%) of the actual F0, and incorrect otherwise. This error criterion corresponds to the smallest rate-pitch discrimination threshold of around 2% for electrical pulse trains obtained by some of the better cochlear implant users (e.g., Moore and Carlyon, 2005). It also captures errors for the smallest F0 intervals encountered in a western musical scale. All tests were conducted on signals presented directly into the audio input of the processor (unless otherwise specified), sampled at 16 kHz in 16-bit resolution, and stored as wave files.

For the sentence material, error rates were measured for sentences spoken by a male and female (Bagshaw et al., 1993) presented directly into the processor. The database can be downloaded from <http://www.cstr.ed.ac.uk/research/projects/fda>. Tests were conducted in quiet and for the same noise conditions as used with the sung
vowels. The F0 error criterion in this case was increased to ±20%, and was assessed relative to laryngograph-derived estimates of F0 provided in the database. The larger criterion was adopted to be consistent with the procedure used by de Cheveigné and Kawahara, (2002) in which a number of alternative F0 estimators were tested using the same sentence material (denoted DB2) as used in the present study. Whilst this error criterion is rather permissive it does capture octave errors which are most problematic for F0 estimators. Furthermore, the sluggishness response and lag of real-time F0 estimators means that errors are inevitable when processing signals in which F0 changes dynamically such as the sentence material tested here. For this material identification of F0 contour (e.g., intonation and voice stress) can be more important than absolute F0 and hence larger errors can conceivably be tolerated without compromising identification of F0 trajectory. A further modification to the error criterion was also used in which F0 estimates were not processed when the broadband (0-2 kHz) STR calculated by the F0 estimator was less than 0.6 (corresponding to a SNR of 1.8 dB). STR values of less than 0.6 reflect a low probability that the signal is monophonic or voiced. In this case channel harmonic probabilities would be close to zero and thus little-to-no F0 modulation would be coded by the strategy and any F0 errors would be of little consequence. Similarly, de Cheveigné and Kawahara, (2002) employed “ground-truth” adjusted F0 estimates\(^{12}\) in their analysis of F0 estimators.

### 4.2.3 STR and harmonic probability analysis

An analysis of both the broadband and channel signal-to-total power ratio estimators was conducted using a complex harmonic tone mixed with white noise or eight-talker babble. Tests were conducted for SNRs ranging from +12 to -12 dB in 4 dB steps. A tone with an F0 of 250 Hz (equal amplitude sine-phase harmonics up to 2 kHz) was selected so that low-order harmonics coincided with centre frequencies of odd numbered BPF bands in the ACE and eTone strategies (the lowest eight BPFs had CFs of 250, 375, 500, 625, 750, 875, 1000, and 1125 Hz). This signal therefore allows comparison of results for channels (1 to 8) that either contain or exclude harmonics of F0. Stimuli were 10

---

\(^{12}\) “Ground-truth” adjusted error rates reported by Cheveigné and Kawahara (2002) were based on adjusted laryngograph estimates of F0 for YIN which rejected estimates that were obviously incorrect and only included those in which vocal fold vibration was evident. Because these adjusted estimates are not available and cannot easily be reproduced we instead included STR qualified error rates which rejected estimates when STR < 0.6 (indicating a lack of periodicity in the signal), thus producing our own form of “ground-truth” F0 estimates. Visual examination of the sentence waveforms and spectrograms confirmed that sections where STR < 0.6 coincided with unvoiced consonants.
seconds in duration and were presented directly into the processor. Results were averaged over the entire duration. Estimated broadband $STR$ and channel $STR_{ch}$ values were compared with the known signal and noise powers in each channel at each test SNR. Channel harmonic probability values $HP_{ch}$ for the complex harmonic signal were also examined to confirm the operation of the harmonic probability estimator.

## 4.3 Results

### 4.3.1 Electrical stimulus patterns

Correct operation of the eTone strategy was verified using electrical stimulus output patterns, known as electrodograms, which are similar to spectrograms for acoustic signals, but plot stimulus intensity for each electrode (channel) as a function of time. In the top-left panel of Fig. 4.3, the electrodogram for ACE processing of the word “choice” spoken by a male speaker with an F0 of approximately 118 Hz is shown. The same word was processed by eTone and is shown in top-right panel where it can be seen that during the voiced vowel (section 2), the stimulus envelope is deeply modulated by eTone at an average frequency equal to the speaker F0. An expanded view of the vowel section processed by each strategy is shown in Fig. 4.4. For the unvoiced consonants (sections 1 and 3), the stimulus signals for eTone and ACE are practically identical. The middle panels display the response of ACE and eTone to one of the sung-vowels used to analyse the performance of the F0 estimator. The vowel was /a/ sung by the male singer at an F0 of 196 Hz. For ACE, the channel envelopes contain varying degrees of F0-modulation with some channels (i.e., electrodes 16, 19 and 20) containing none at all. In contrast, the response for eTone shows deep F0 modulation in most channels that is coincident in time across channels. The sung-vowel was also presented in eight-talker babble noise at a SNR of +6 dB and the response for ACE and eTone to this signal are shown in the bottom panels of Fig. 4.3.
Chapter 4: Development of a Rate-Pitch Strategy (eTone)

Figure 4.3. Electrical stimulus patterns (electrodegram) recordings for the word “choice” processed by the ACE and eTone strategies in panels (a) and (b) respectively, and for a shortened (280 ms) sung-vowel /a/ (F0 = 196 Hz) presented in quiet and in eight-talker babble noise at an SNR of +6 dB for ACE in panels (c) and (e) respectively and for eTone in panel (d) and (f) respectively. Time is plotted along the abscissa and electrode number along the ordinate. For each electrical pulse, a vertical bar is shown at the time and electrode site of stimulation. The height of each bar represents the stimulus intensity (in clinical units or log current). For both strategies, 10 maxima from 22 channels were selected. The adaptive $K_p$ algorithm was used for eTone with the quiet processing mode enabled in panels (b) and (d), and the noise processing mode enabled in panel (f).
For eTone, F0 modulation is still coded on most channels pertaining to the first and second formants of the vowel, thus demonstrating the robustness of the F0 estimator to noise. The fact that the modulation is not seen in channels containing predominantly noise further verifies the correct operation of the channel harmonic probability estimator. However, careful examination also shows that the coded modulation frequency is occasionally in error and the depth of modulation varies with the channel harmonic probability. In addition, for both strategies, some of the weaker formant energy (i.e., on electrodes 21, 22, 14, 7, 8, and 9) are not coded due to selection of higher amplitude maxima in noise channels. For eTone, these channels have a low harmonic probability and thus little F0 modulation is coded. A final point of note is that the spectral information (i.e., channel number and stimulus intensity) and low-frequency (< 50 Hz) temporal envelope information (i.e., excluding F0) coded by strategies to each stimulus in Fig. 4.3 are very similar. It is not unreasonable to thus assume that these strategies may provide similar levels of performance in speech perception by CI users who utilise spectral and low-frequency temporal envelope information predominantly to discriminate phonemes.

4.3.2 F0 estimator analysis

F0-estimator error rates averaged across the two sung vowels (/a/ and /i/) using a ±0.5 semitones accuracy criterion are plotted as a function of input SNR in Fig. 4.5 (a) and (b) for the male and female stimuli, respectively. Error rates in quiet using the quiet processing mode and a fixed value of $K_W$ were 0.86%, 0.13%, and 0.66%, for the male, female, and average of both singers’ stimuli, respectively. Using adaptive $K_W$, error rates in quiet were almost identical to those when $K_W$ was fixed (0.84%, 0.13%, and 0.65%, respectively). At an SNR of 0 dB using the noise processing mode and a fixed value of
Chapter 4: Development of a Rate-Pitch Strategy (eTone)

$K_W$, error rates for the same stimuli were 8.18%, 0.13%, and 6.03%, respectively for white noise, and 23.43%, 8.06%, and 19.33%, respectively for babble noise. Using adaptive $K_W$, error rates for white noise were slightly lower (3.37%, 0.18%, and 2.52%), whereas for the babble they remained largely unchanged (23.91%, 11.0%, and 20.47%). The results for adaptive $K_W$ were always close to the best-of the fixed $K_W$ values demonstrating that the adaptive algorithm operated well, at least for the set of stimuli and noise conditions tested here.

![Figure 4.5. F0 estimation error rates as a function of input SNR for the sung-vowel male and female stimuli in panels (a) and (b) respectively. Results for white noise are denoted by triangles and those for the babble are denoted by circles. For each noise type, error rates are plotted for the quiet (dashed-lines) and noise (dotted lines) processing modes using fixed values of $K_W$ as well as for the adaptive $K_W$ algorithm (solid lines). For the quiet processing mode, $K_W = 0.08$ and the frame buffer size = 4, whereas for the noise processing mode, $K_W = 0.3$ and frame buffer size = 12. For the adaptive algorithm, $K_W$ varied between 0.02 and 0.34 and the frame buffer size = 4 for SNRs $\geq +8$ dB, otherwise it was increased to 12 (using the processing mode switch). All other processing parameters were fixed between conditions (as specified in the methods). RMS signal and noise levels were measured using a flat response across the wide-band (8 kHz) frequency range. If instead levels were measured using A-weighting, then SNRs for the sung vowels in babble noise would be approximately 1 dB lower than wide-band SNRs (i.e., 0 dB = -1 dBA), whereas SNR levels for the sung vowels in white noise would be approximately 2 dB higher.]

The measured error rates were substantially lower than those reported by Zakis et al., (2007) using an earlier version of the F0 estimator. Using a subset of the sung-vowel stimuli (i.e., vowel /a/ only), those authors reported error rates for the male, female, and average of both singers’ of 4.5%, 2.0%, and 4.1% respectively in quiet, approximately 11%, 2%, and 8% respectively for white noise at 0 dB SNR, and approximately 45%, 22%, and 38% respectively for babble noise at 0 dB SNR. In the present study, using
Chapter 4: Development of a Rate-Pitch Strategy (eTone)

the same subset of stimuli as Zakis et al., (2007), adaptive $K_w$ error rates for the male, female, and average of both singers’ were 0.0%, 0.0%, and 0.0% respectively in quiet, 1.2%, 0.13%, and 0.9% respectively for white noise at 0 dB SNR, and 20.0%, 8.5%, and 16.9% respectively for babble noise at 0 dB SNR.

Error rates were also examined for stimuli presented in the sound-field at 65 dB SPL (measured using A-weighted fast response average metering). A subset of the sung-vowel stimuli (i.e., male vowel /a/ spaced by three semitone intervals), were examined. Using fixed values of $K_w$ and the adaptive $K_w$ algorithm, similar error rates in quiet to those for direct input were obtained. However with the addition of noise, error rates in the sound field increased compared to those for direct input by 0.9 and 1.9 percentage points for white noise and babble noise respectively at an SNR of +8 dB, by 3.4 and 2.2 percentage points respectively at an SNR of +4 dB, and by 6.5 and 7.0 percentage points respectively at an SNR of 0 dB. The increased error rates in noise were attributed to room effects (i.e., reverberation) and the addition of processor and microphone noise particularly at low frequencies which had been amplified relative to higher frequencies by the -4 dB/octave de-emphasis applied for sound field input.

Figure 4.6. F0 estimation error rates using the adaptive $K_w$ algorithm as a function of input SNR for the sentences spoken by the male and female speakers in panels (a) and (b) respectively. Results for white noise are denoted by triangles and those for the babble are denoted by circles. Error rates are plotted for the standard F0 ±20% error criterion (dashed lines and open symbols) and for the $STR > 0.6$ error criterion (solid lines and filled symbols). For SNRs ≥ +8 dB, the frame buffer size = 4, otherwise it was increased to 12 (using the processing mode switch). Although not shown, adaptive $K_w$ error rates were as good as the best-of quiet or noise processing mode fixed $K_w$ parameters. Using A-weighting of levels, SNRs for the sentences in babble noise were unchanged and those for the sentences in white noise were approximately 1 dB higher (i.e., 0 dB = 1 dB$_A$).
Chapter 4: Development of a Rate-Pitch Strategy (eTone)

For the sentence material, error rates using the adaptive $K_W$ algorithm are plotted as a function of input SNR in Fig. 4.6 (a) and (b) for the male and female speakers respectively. Error rates in this case are shown for the “standard” F0-accuracy criterion of $\pm 20\%$ (dashed lines) and for the $STR > 0.6$ criterion (solid lines). Error rates in quiet were 3.98%, 2.38%, and 3.13%, for the standard criterion and 1.36%, 1.64%, and 1.51% for the $STR > 0.6$ criterion, for the male, female, and average of both speakers’ stimuli respectively. These rates compare favourably to “ground-truth” adjusted error rates reported by de Cheveigné and Kawahara (2002) who evaluated a number of different F0 estimators using the same recorded sentences (see Table II, column 3: DB2). The average error rate for the eleven algorithms evaluated in that study was 8.4% (averaged across the male and female DB2 sentences) and the lowest error rate was 1.4% using the YIN algorithm, which is comparable to the error rate of 3.13% and the $STR$ qualified error rate of 1.51% for the current F0 estimator. With the inclusion of noise, error rates using the $STR$ qualified criterion remained below 10% for SNRs as low as -4 to -8 dB for white noise, and +4 dB for babble noise.

Overall across all tests and noise conditions using the adaptive $K_W$ algorithm, error rates were below 2% in quiet conditions and remained below 10% for SNRs up to approximately +4 dB. Whilst these error rates were low, they could impact on perception of F0 by cochlear implant users. The errors often manifested as octave errors spanning a short portion (i.e., 10-40 ms) of the stimulus. Given that average F0 rate difference limens (DLs) are around 7% for reference rates of 100 Hz using sequentially presented electrical pulse trains (e.g., Moore and Carlyon, 2005), it is possible that short-duration changes in rate of an octave could alter pitch. Stohl et al., (2009) observed that instantaneous rate changes of 100% or higher in electrical pulse trains could be detected, but not necessarily discriminated, when their duration was greater than approximately four cycles of the reference period (i.e., 20 ms for 200 Hz). However, considerable variability was observed when duration was varied adaptively. Other studies have shown that whilst rate-pitch can be discriminated for durations as short as 20ms, DLs generally improve for durations up to approximately 100 ms, supporting a multiple-look model in which a critical duration of five-ten cycles may be needed to perceive changes in pitch (e.g., Lou et al., 2010). Taken together, those data suggest that while rate changes due to octave errors might be detected, it is not clear whether they would introduce a noticeable change in pitch given the short-duration of many of those errors.
4.3.3 STR and harmonic probability analysis

Broadband $STR$ and channel $STR_{ch}$ (for channels 1 to 8) are plotted in Fig. 4.7 as a function of input SNR in panels (a) and (b) for white noise, and babble noise, respectively.

![Figure 4.7](image)

Figure 4.7. Broadband $STR$ and channel $STR_{ch}$ for channels 1 to 8 plotted as a function of input SNR for (a) white noise, and (b) babble noise. Broadband $STR$ is plotted as solid lines using cross symbols and channel $STR_{ch}$ as dashed lines. Panels (c) and (d) plot $STR$ and channel $STR_{ch}$ accuracy as estimated/actual $STR$ for white noise and babble noise respectively. Panels (e) and (f) plot channel harmonic probability $HP_{ch}$ for white noise and babble noise respectively. Odd number channels which are centred at harmonic frequencies of the test signal are plotted with filled symbols whereas as even numbed channels that contain little harmonic energy are plotted with open symbols. Using A-weighting for the test signal, input SNRs for white noise are unchanged whereas those for babble noise are approximately 1 dB lower than wide-band SNRs.

Estimation accuracy, plotted as the ratio of estimated over actual $STR$, is shown in panels (c) and (d) for white noise, and babble noise, respectively. F0 estimator broadband $STR$ accuracy remains between 0.5 and 2.0 (or $\pm 3$ dB compared to unity) for
input SNRs of up to approximately -8 dB for white noise and -4 dB for babble noise. At lower SNRs, it is limited by the ratio of sieve bandwidth $S_{BW}$ to total bandwidth $T_{BW}$ for each sieve. Similarly, for channels containing harmonics of F0 (i.e., Chs 1, 3, 5, and 7), channel $STR_{ch}$ accuracy remains between 0.5 and 2.0 for SNRs up to approximately -12 dB for white noise, and 0 dB for babble noise. For more adverse noise conditions, channel $STR_{ch}$ is generally overestimated. In contrast, for channels not centred at harmonic frequencies (i.e., Chs 2, 4, 6, and 8), $STR_{ch}$ accuracy rapidly deteriorates beyond 2.0 for SNRs less than or equal to approximately +8 dB for white noise, and +12 dB for babble noise. This occurs because these channels contain very little F0 energy and thus any noise in these channels will dominate the estimated $STR_{ch}$ value. However for these cases, poor channel $STR_{ch}$ accuracy will be of little practical consequence because these channels have a low $STR_{ch}$ and thus their harmonic probability $HP_{ch}$, as shown in panels (e) and (f) for white noise, and babble noise, respectively, will also be low resulting in coding of little-to-no F0 modulation. In addition, these channels contain less energy than those with both signal and noise components, and are therefore less likely to be selected as maxima by the eTone strategy. Correct operation of the channel harmonic probability estimator can be observed in panels (e) and (f). For channels containing harmonics of F0, $HP_{ch}$ values remain close to 1.0 for SNRs up to 0 dB for white noise, and +4 dB for babble noise. In contrast for channels excluding harmonics of F0, $HP_{ch}$ values decrease rapidly for SNRs below approximately +12 dB for white noise, and +8 dB for babble noise.

Similar performance to that seen above was observed with sung-vowel stimuli. However, because the harmonic content of channels varies with each stimulus, presentation of averaged outcomes is uninformative. Similar outcomes to those shown above were also observed for channel $STR_{ch}$ derived from the autocorrelation method (i.e., for channels with CFs above 2 kHz). However in this case, channel $STR_{ch}$ accuracy was less robust to the effects of noise and, as such, a lower inflection point was employed in the sigmoid function ($a = 0.35$ in Eq. 4.13) to derive the channel harmonic probabilities. For these channels, harmonic probabilities were more dependent on the broadband $STR$ rather than channel $STR_{ch}$.

### 4.4 Discussion

Laboratory analysis of the experimental eTone strategy established that it operates as per the design requirements described in the introduction. The F0 estimator provides a
quick and reliable estimate of F0 that is robust to the effects of noise and therefore allows accurate F0 information to be encoded in the stimulus envelope. Accuracy of estimated signal-to-total power ratio for the F0 estimator signal and for each BPF channel signal was confirmed for input SNRs up to approximately 0 dB (which covers the range of SNRs commonly encountered by users of CIs). Thus reliable predictions of channel harmonic probabilities and appropriate scaling of F0-modulation depth in each channel can be achieved. The F0 modulation waveform coded in eTone channels produced much more distinct F0 timing information in the stimulus envelope than ACE. However, whether or not this processing produces pitch percepts that are also more salient, accurate, and robust to the effects of noise than that of ACE can only be investigated using perceptual tests. A study with six CI users has been carried out and significant benefits compared to ACE were observed in pitch ranking tests. Those outcomes are presented in the next chapter.

An unavoidable consequence of the processing used in eTone is that it introduces an additional delay in the signal path. The standard ACE strategy has a processing delay equal to half the FFT window duration (i.e., 4 ms). The F0 estimator and envelope tracker stages in eTone introduce an additional delay of 13.75 ms. Previous studies (e.g., Vandali, 2001, Vandali et al., 2005) have shown that delays on this order are not noticed by the majority of CI users. Some however do notice this delay but they quickly become accustomed to it. Stone and Moore (2002) observed that for speech production, delays between 25 and 30 ms were disturbing. Furthermore for speech perception where lip synchronisation is used, delays of around 15-20 ms can become disturbing. The total delay of 17.75 ms is thus just on the border of being unacceptable. Unfortunately, not much can be done to reduce this delay in the present implementation without compromising performance of the strategy although it is possible that shorter delays are achievable using alternate envelope tracking techniques.

F0-modulated channel signals were produced via amplitude modulation of the slow-varying channel envelope signals. The slow-varying channel envelopes followed peak levels in the standard ACE channel envelope signals. Thus, the F0 modulated channel signals maintained the same peak level as the standard envelope signals but incorporated deep F0 modulation. The loudness of these stimuli might therefore be lower with eTone than ACE (§2.3.3.6), as was shown for similar stimuli in chapter 3. Just how much lower is likely to be subject-specific. The modulation gain constant \( K_{HP} \) was included for that reason to provide additional gain to F0-modulated channel
signals. Preliminary tests in a group of six subjects showed preferred gains ranging from +1 to +6 dB. Alternative methods of compensating for differences in loudness include adjusting the range of electrical stimulation levels for F0 modulated channel signals, rather than the gain, or adaptively measuring the RMS level, or the specific loudness (Moore and Glasberg, 1996), of the channel envelope signal and reproducing the same level or loudness for the F0-modulated signal.

A simple optional method of reducing the contribution of noise in the estimated harmonic signal power was described in stage four of the F0 estimator. With omission of this stage a small reduction in F0 estimation accuracy and slightly overestimated broadband STR is observed for moderate-to-low SNRs. However, the method assumes that the signal is monophonic and that noise is uniformly distributed over the 0-2 kHz frequency range and thus its performance benefit may not be maintained for different noise types, such as narrow-band noise or for polyphonic signals. Whilst not tested, this assumption could be eased by subdividing the total bandwidth into a number of smaller frequency bands (e.g., of bandwidth equal to the sieve F0 and centred around each harmonic band). Harmonic signal powers within each narrow-band could then be calculated from Eq. (4.10) (using $K_{BW} = 1.0$) and summed across bands to give a better estimate of the matched harmonic signal power. Further improvements may be available through the use of front-end noise reduction algorithms, such as beamformers (e.g., Wouters and Vanden Berghe, 2001).
5 Assessment of the eTone Strategy

The abilities to hear changes in pitch for sung vowels and understand speech using an experimental sound coding strategy (eTone described in chapter 4) that enhanced coding of temporal fundamental frequency (F0) information were tested in six cochlear implant users, and compared with performance using their clinical (ACE) strategy. In addition, rate- and modulation rate-pitch difference limens (DL) were measured using synthetic stimuli with F0s below 300 Hz to determine psychophysical abilities of each subject and to provide experience in attending to rate cues for the judgment of pitch. Sung-vowel pitch ranking tests for stimuli separated by three semitones presented across an F0 range of one octave (139 Hz to 277 Hz) showed a significant benefit for the experimental strategy compared to ACE. Average d-prime (d’) values for eTone (d’=1.05) were approximately three time larger than for ACE (d’=0.35). Similar scores for both strategies in the speech recognition tests showed that coding of segmental speech information by the experimental strategy was not degraded. Average F0 DLs were consistent with results from previous studies and for all subjects were less than or equal to approximately three semitones for F0s of 125 and 200 Hz.

5.1 Introduction

Cochlear implant (CI) systems have been shown to be clinically effective in providing a sensation of hearing and varying degrees of open-set speech recognition to individuals with a severe to profound hearing loss (§2.1). However, perception of musical and voice pitch has been shown to be far from satisfactory, as demonstrated in studies investigating discrimination of musical pitch intervals (§2.4.1) and lexical contrasts in tonal languages (§2.5).

In normal hearing (NH), pitch is primarily encoded via fine spatio-temporal structure for signals that can be resolved by auditory filters responsive to low order harmonics of the fundamental frequency (F0) (§2.2.1). Pitch can also be encoded in the temporal envelope of wider auditory filters responsive to higher order unresolved harmonics, for which F0 beating results from the interaction of those harmonics. However, pitch perception derived from envelope information has been shown to be poorer and more sluggish than that derived from fine spatio-temporal structure (§2.3.3.2, Shackleton and Carlyon, 1994; Plack and Carlyon, 1995; White and Plack, 1998).
To date, little evidence has been forthcoming that fine temporal structure beyond approximately 300 Hz, let alone fine spatio-temporal structure, can be conveyed by present CI devices despite various attempts to do so (§2.6.4). Psychophysical studies have demonstrated that rate DLs for electrical pulse trains typically deteriorate markedly above 300 pulses-per-second (PPS) (2.3.2), as well for modulation rates above 300 Hz with electrical sinusoidal amplitude-modulated stimuli (§2.3.3). In addition, studies examining the spread of neural excitation along the cochlea for electrical stimuli (§2.3.1.2) have shown that the spatial excitation patterns from adjacent electrodes can overlap significantly. Such overlap limits the encoding of independent temporal information at nearby places that would be needed to represent fine spatio-temporal information along the cochlea, as has also been borne out in psychophysical studies examining integration of temporal information as a function of electrode separation (McKay and McDermott, 1996; Macherey and Carlyon, 2010). Both those limitations (temporal and spatial) impede the ability to code fine structure in present CI devices as was seen for example in Vandali et al., (2005) in which the Peak Derived Timing (PDT) strategy, which codes fine temporal detail by generating stimuli that correspond in time and amplitude to positive temporal peaks of each band-pass filtered channel, provided no significant benefit to pitch perception when compared to envelope based coding strategies.

It seems unlikely therefore that the fine temporal and/or spatio-temporal pitch encoding mechanism used in NH can be exploited using present CI devices. However, the envelope-based rate-pitch mechanism, which is thought to be similar to that of rate and modulation-rate-pitch in electrical hearing (McKay and Carlyon, 1999), can provide useful, albeit poorer performance. For instance, in a review of five studies, Moore and Carlyon (2005) reported an average electrical rate-pitch DL of 7.3% (with a range of approximately 2 to 18%) for a rate of 100 Hz (§2.3.2.2). While that value is substantially worse than those observed for complex tones in normal hearing (e.g., Moore and Peters, 1992), it suggests that some CI users may be able to rank pitch intervals approaching one semitone (5.9%), at least at low F0 rates. Another limitation of envelope-derived pitch relates to its sluggish temporal response. Lou et al., (2010) examined the effects of stimulus duration on modulation frequency discrimination in CIs users. Their findings suggested that a critical duration of 5-10 modulation cycles may be needed by CI users to discriminate modulation frequency (§2.3.3.2). This result is similar to that observed in NH for unresolved harmonic stimuli (Plack and Carlyon,
Chapter 5: Assessment of the eTone Strategy

1995; White and Plack, 1998) and is in contrast to results with resolved harmonic stimuli that demonstrate a relatively shorter integration time.

Previous studies examining pitch perception using experimental CI strategies that code rate-pitch information in the stimulus envelope, have demonstrated that significant benefits can be obtained compared to clinical strategies (e.g., Green et al., 2004, 2005; Vandali et al., 2005; Laneau et al., 2006) by providing deeper F0 modulation that is in phase across channels (§2.6.3). That approach provides more salient cues to F0 and reduces unwanted temporal interactions arising from the spread of the excitation field.

One such strategy, known as Multi-channel Envelope Modulation (MEM) (Vandali et al., 2005; §2.6.3.3), was also evaluated using lexical-tone recognition tests in a group of nine Cantonese speaking CI users (Ciocca et al., 2005). However, results failed to demonstrate any significant benefits using MEM as compared to the Advanced Combinational Encoder (ACE) strategy (Vandali et al., 2000) or the Continuous Interleaved Sampling (CIS) strategy (Wilson et al., 1991). Possible explanations of that outcome include: a) the F0 differences between tonal contrasts (which can be as low as one or two semitones) were too small to be discriminated by those CI users evaluated who obtained average F0-rate DLs of approximately 2.5 semitones for a 160-Hz three-harmonic complex tone; b) the rate-pitch mechanism exploited in CIs is too sluggish to be useful in discrimination of F0 contours in Cantonese tones, which can have short durations and can undergo rapid changes in F0 (e.g., Xu, 1997); c) listeners may have been unaccustomed to attending to rate information for judging F0/tone because those cues are poorly coded by their clinical strategy (e.g., Vandali et al., 2005) and so they attend other cues in everyday listening, such as intensity, duration, and context in running speech for discrimination of tone; and d) conflicting cues to pitch coded by place of stimulation may have disrupted temporal rate-pitch information (Laneau et al., 2004b; Vandali et al., 2005). One of the goals of the present work was to reduce the effects of the some of the above problems, namely points a), and c), by conducting experiments that directed listeners’ attention to rate cues when judging pitch.

Another explanation for the results of Ciocca et al., (2005) is that the coded rate-pitch information was not salient enough and/or may not have accurately reflected F0. To address that problem a new envelope-based rate-pitch coding strategy, Enhanced-Envelope-Encoded Tone (“eTone”), has been formulated to encode rate-pitch information in a more salient and accurate manner and to be more robust to the effects of noise than those previous experimental and present clinical strategies (see chapter 4).
Chapter 5: Assessment of the eTone Strategy

The strategy is based on ACE and includes additional processing that explicitly applies synchronous F0 modulation to channels that contain harmonics of a prominent complex tone. Deep modulation with a sharp temporal onset is employed so that F0 is well represented in neural response timing (e.g., similar to Green et al., 2004, 2005; §2.6.3.2). Channels that contain only inharmonic signals retain envelopes normally produced by ACE. A novel real-time F0 estimator is employed to track F0 to within ±0.5 semitones. Robustness to the effects of noise for signal-to-noise ratios that CI users are typically able to accommodate (i.e., down to approximately 0 dB) was verified.

In the present study, performance with both the experimental strategy and ACE was examined using pitch ranking tests in a group of six cochlear implant users. Pitch ranking was measured for sung-vowel stimuli separated by three semitones and presented in the sound field. Previous unpublished work by the authors had indicated that variations in the number of maxima (the number of channels with the largest amplitudes selected for stimulation) might affect a listener’s judgment of pitch, particularly for musically inexperienced listeners attending largely to changes in place of stimulation to judge pitch. For instance a reduction in the number of maxima selected can impart small changes in coded place information due to omission of some spectral energy thereby altering the pitch of the sound. For that reason, two variations of the number of spectral maxima were examined for each strategy. To address some of the issues discussed earlier, a series of rate- and modulation rate-pitch DL tests were conducted to direct subjects’ attention to the use of rate cues for judgment of pitch, and to establish the range of F0 intervals that could be discriminated by each subject when presenting synthetic “idealised” stimuli through a sound processor. The sung-vowel pitch-ranking tests were conducted both before and after the pitch DL measures. Speech perception tests in quiet, and in noise, were also administered to ensure that the processing employed by eTone was not detrimental to speech recognition.

5.2 General methods

5.2.1 Subjects

Six post-lingually deaf adult users of the Nucleus 24 cochlear implant system took part in this study. Their demographic details are listed in Table 5.1. S1 was the only subject who had formal music training prior to the onset of deafness. S5 was a speaker of tonal
languages and was thus experienced in tone-discrimination. All subjects indicated that they had enjoyed listening to music prior to deafness.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Gender</th>
<th>Age (yr)</th>
<th>Years implanted</th>
<th>Musical/Tonal experience</th>
<th>Cochlear Implant</th>
<th>Clinical Processor</th>
<th>Clinical Strategy</th>
<th>Stimulation rate (PPSCH)</th>
<th>Channels</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>M</td>
<td>59</td>
<td>1</td>
<td>Yes</td>
<td>CI24RE(CA)</td>
<td>Freedom</td>
<td>ACE</td>
<td>900</td>
<td>22</td>
</tr>
<tr>
<td>S2</td>
<td>M</td>
<td>72</td>
<td>10</td>
<td>No</td>
<td>CI24M</td>
<td>Freedom</td>
<td>SPEAK</td>
<td>250</td>
<td>20</td>
</tr>
<tr>
<td>S3</td>
<td>F</td>
<td>86</td>
<td>7</td>
<td>No</td>
<td>CI24R(CS)</td>
<td>ESPrit3G</td>
<td>SPEAK</td>
<td>250</td>
<td>20</td>
</tr>
<tr>
<td>S4</td>
<td>F</td>
<td>58</td>
<td>7</td>
<td>No</td>
<td>CI24R(CS)</td>
<td>ESPrit3G</td>
<td>ACE</td>
<td>1200</td>
<td>20</td>
</tr>
<tr>
<td>S5</td>
<td>M</td>
<td>53</td>
<td>10</td>
<td>Yes</td>
<td>CI24M</td>
<td>ESPrit3G</td>
<td>ACE</td>
<td>900</td>
<td>20</td>
</tr>
<tr>
<td>S6</td>
<td>F</td>
<td>61</td>
<td>6</td>
<td>No</td>
<td>CI24R(CS)</td>
<td>ESPrit3G</td>
<td>ACE</td>
<td>900</td>
<td>20</td>
</tr>
</tbody>
</table>

Table 5.1. Subject details: gender, age, number of years implanted at commencement of the study, musical/tonal-language experience prior to deafness onset, implanted cochlear implant device (CS = Contour System, CA = Contour Advance; otherwise straight array), clinical speech processor and coding strategy, stimulation frequency per channel, and number of channels employed in clinical map.

5.2.2 Strategies

Both the Enhanced-Envelope-Encoded Tone (eTone) strategy and ACE were implemented in the SPEAR3 research sound processor. The implementation of ACE evaluated in this study was similar to that used in clinical devices, but did not include any of the pre-processing stages such as automatic sensitivity control (ASC), automatic gain control (AGC), or adaptive dynamic range optimisation (ADRO), all of which could introduce uncontrolled gain variations during testing. The eTone strategy is based on ACE but includes additional processing to code F0 modulation in the stimulus envelope of each channel, refer to chapter 4 for a complete description of eTone. For complex tones (e.g., a voiced vowel or musical sound played by a tonal instrument), the slow-varying (≤ 70 Hz) envelope of the stimulus signal in each channel is modulated at the F0 rate. The modulation function is similar to the EDM function described in chapter 3. It consists of a narrow-pulse with unity peak-level and an exponentially decay time set to approximately 25% of the modulation period. The slow-varying envelope signal in each channel is modulated by this function such that its peak level is preserved and its modulation depth (peak-to-trough) is approximately equal to half that of the electrical dynamic range (DR) coded on each electrode. That depth is approximately equivalent to 20 dB in the acoustic signal path given an input DR of 40 dB. For noise or inharmonic signals (e.g., an unvoiced consonant or percussive sound),
Chapter 5: Assessment of the eTone Strategy

the electrical stimulus is derived from each channel’s narrow-band envelope response in the same manner as ACE. For signals that simultaneously comprise harmonic and inharmonic components, a mixture of the F0-modulated and the non-modulated envelope signals is coded in each channel. The mixing ratio for the two envelopes is derived from a channel harmonic probability estimator which determines the degree to which each channel signal is related to the most dominant F0 in the signal. A real-time F0 estimator that operates over an F0 range of 80 to 400 Hz is used to determine F0.

5.2.3 Strategy fitting

The eTone and ACE strategies evaluated in the present study employed a stimulation rate of 1455 pulses-per-second/channel (PPS/CH), which allowed F0 modulations up to approximately 360 Hz to be coded (§2.3.3.5). Threshold and maximum comfortable loudness levels of stimulation were thus measured at a rate of 1455 PPS for each electrode used in each subject’s clinical map. Those levels were entered into the ACE map and globally adjusted, if necessary, for clarity and comfort when listening to speech. The same levels were then entered into the eTone map. Two map variations were generated for each strategy, one employing eight maxima which is the preferred clinical value, and the other using ten which was chosen to reduce omission of spectral place information that could otherwise adversely affect a listener’s judgment of pitch. In addition, two parameters of eTone (modulation gain and depth) were adjusted for each subject to compensate for any loudness and speech clarity differences between eTone and ACE (refer to results of experiment 3 in §3.4 and §3.6.2). Loudness differences could arise for harmonic signals due to differences in the modulation waveform duty cycle, which is generally much higher for ACE than for eTone (§2.3.3.6). For equal peak envelope levels, loudness is therefore expected to be higher using ACE. For each subject, loudness for eTone was adjusted using a gain constant applied only to the F0 modulated component of the stimulus signal. Approximate loudness matching between strategies when listening to spoken vowels, /a/ and /i/ was achieved using gain constants of 6 dB for S1, 5 dB for S3, 4 dB for S4 and S5, 3 dB for S2, and 1 dB for S6. The modulation depth parameter was typically set to 50% of the electrical dynamic range (which is equivalent to approximately 20 dB in the acoustic signal path), except for S4 for whom it was set to 30% (12 dB) to reduce reported harshness and improve speech clarity. All other strategy parameters, excluding those of the pre-processing stages discussed earlier, were programmed identically to the subject’s clinical map. After
fitting of the strategies, subjects commented that “quality” or “clarity” of speech was very similar between strategies, but also reported that voices sounded deeper when using eTone compared to ACE. Two subjects, S4 and S5, noticed the additional processing delay of 13.75 ms introduced by eTone when listening to their own voice and described it as an “echo”. After using eTone for a short time, S5 commented that the echo was not bothersome, whereas S4 indicated that it still distracted her and would take some time to adjust to.

5.2.4 Experimental protocol

The study consisted of three different experiments. In experiment 1, pitch DLs were measured using five pitch coding techniques/conditions. The first coded pitch using unmodulated pulse rate on a single electrode. The others conveyed pitch through modulation rate and were based on modified versions of ACE (which provides sinusoid-like amplitude modulation) and eTone (sharp-onset exponential-decay EDM modulation) that restricted stimulation to one, or five, adjacent electrodes. Complex harmonic tone stimuli were provided as input to those modified strategies and are described further in §5.3.1. In experiment 2, pitch ranking performance was measured using ACE and eTone for sung-vowel stimuli separated by three semitones. For this case, place of stimulation for each stimulus could vary with F0 due to slight differences in the spectral envelope of the stimuli across F0. In experiment 3, speech perception was assessed in quiet, and noise, using ACE and eTone. The study was conducted over ten two-hour sessions, conducted once per week. In the first session, the strategies were fitted in the SPEAR3 processor. No take home experience with either strategy was provided. In session two, pitch ranking tests (experiment 2) were conducted. In session three, the speech perception tests (experiment 3) were administered. In sessions four to eight, rate- and modulation rate-pitch DLs (experiment 1) were measured. Sessions four to eight also served to encourage/direct subjects to utilise temporal cues to judge pitch in the absence of electrode-place variations. In session nine, the pitch ranking tests were repeated to assess whether any benefits were gained from subjects’ added experience in using rate-cues to judge pitch in experiment 1. However, the absence of a control group meant that any change across sessions could not be directly attributed to the effects of the intervening pitch DL tests of experiment 1. In the final session, speech perception tests were repeated. Subject S6 withdrew from the study mid-way through experiment 1 (i.e., at session seven) due to a medical condition that was unrelated to her hearing
prosthesis. Some of the modulation-rate DLs were therefore not collected for that subject and the second batch of pitch ranking and speech perception tests were not administered.

5.2.5 Data Analysis

In all tests, group averaged data was subjected to analysis of variance (ANOVA) using a general linear model in which subject was treated as random (blocked) factor followed by post-hoc Bonferroni tests ($p < 0.05$) to determine any specific treatment differences. Prior to the analysis, normality of the data was tested using either a Shapiro-Wilks test or by confirming that the probability distribution of residuals remained within 95% confidence limits of nominal values. Data were also checked for homogeneity of covariance using Mauchly’s test of sphericity or by examining the distribution of residuals versus fitted values. When evidence of non-normally distributed data or the assumption of homogeneity was not fulfilled, a non-parametric ANOVA was performed.

5.3 Exp. 1: Stimulation rate- and modulation rate-pitch difference limens

5.3.1 Methods

Unmodulated rate-pitch DLs were measured at three reference rates of 125, 200, and 275 PPS, for electrical pulse trains presented at a single electrode site (fifth most apical electrode corresponding to channel 5 in each subject’s map) using a monopolar electrode configuration and biphasic pulses of 25 μs duration with an 8 μs inter-phase gap. The unmodulated “PulseTrain” stimuli (first coding condition in Exp. 1) were generated using custom software on a PC computer which interfaced to the SPEAR3 processor. The test employed a two-interval, two-alternative, forced choice (2I-2AFC) procedure in which subjects were asked to nominate which of two randomly ordered stimuli differing in rate was higher in pitch. The stimulus pairs differed in rate by 0.5, 1, 2, 3, 4, or 5 semitones around the reference rate. All pairs were presented 10 times in randomised order of rate and interval within the same block. For each reference rate, at least two blocks were administered. The loudness of the stimuli was set to a comfortable level for each subject and loudness balanced across rates using a 1-up 1-down staircase procedure (average of last four turning points). The intensity of each stimulus was roved by up to approximately -10% of each subject’s electrical DR (in clinical units) to reduce any
systematic use of loudness when judging pitch. A 300-ms stimulus duration and 300-ms inter-stimulus interval was used.

On each trial, subjects were instructed to indicate which of the two stimuli was higher in pitch and to ignore any variations in loudness. If unsure, they were told to provide their best guess. No repetition of the stimuli was allowed and no feedback was provided. At commencement of the experiment, if a subject was not familiar with the concept of pitch, he/she was told that pitch is the quality of sound that distinguishes for instance two different notes played on a musical instrument such as a saxophone or the voice of a male talker relative to a female or child’s voice. In some cases subjects listened through their speech processor to an acoustically presented vowel, /a/ at a low F0 of approximately 110 Hz, followed by the same vowel presented approximately one octave higher and was told that the second was higher in pitch. Response bias-corrected percent-correct scores (Pcmax) were used to determine discrimination thresholds for a performance criterion Pcmax = 76% as per the methods described in section 3.3.1.1.

In addition to the PulseTrain coding method, DLs for four additional conditions were obtained. Modulation rate-pitch DLs were measured at reference modulation rates of 125, 200, and 275 Hz using complex tones presented directly to the SPEAR3 processor’s external input and processed using modified versions of ACE and eTone that ensured that electrode-place of stimulation remained fixed. The complex tones were generated by summing equal-amplitude, sine-phase harmonics of F0 up to 1400 Hz, and subsequently applying a 4th order band-pass filter (BPF) with -3 dB corner frequencies at 550 and 1100 Hz. All tones were RMS level balanced. The same 2I-2AFC test procedure, instructions, and method of threshold determination as described for unmodulated rate-pitch DLs were used except in this case levels were roved by up to -6 dB, corresponding to approximately 15% of the available electrical DR. The four additional coding conditions examined included 2 strategy variations by 2 numbers of activated electrodes. The first (denoted ACE1) processed the signal using a modified ACE strategy that employed a single channel/electrode map. The same electrode as used in the unmodulated rate-pitch experiment was activated. The single channel BPF corner frequencies were 300 and 1400 Hz so that envelope fluctuations above the highest modulation frequency examined were preserved. The resultant electrical stimuli consisted of amplitude-modulated pulse trains presented at a single site with smooth modulation functions that were sinusoid-like in terms of clinical levels (i.e., log current),
Chapter 5: Assessment of the eTone Strategy

and with modulation depths in excess of 50% of the electrical DR, (see top panel of Fig. 5.1).

Figure. 5.1. Example electrical stimulation patterns for the amplitude-modulated stimuli presented in experiment 1. F0 for the complex harmonic input signal was 205.9 Hz (i.e., +0.5 semitones above a 200 Hz reference rate). From top to bottom panel, stimulus patterns for the ACE1, eTone1, ACE5, and eTone5 coding conditions are plotted respectively. Time is plotted along the abscissa and electrode number along the ordinate. For each electrical pulse, a vertical bar is shown at the time and electrode site of stimulation.

The next coding condition (denoted eTone1 shown in the second panel of Fig. 5.1) used the eTone strategy with a single electrode map, again using the fifth most apical electrode. The resultant modulation waveform had a sharp-onset and a rapid-exponential decay, and the modulation depth relative to its peak level was equal to 50% of the DR. The final two coding conditions (denoted ACE5 and eTone5) increased the number of activated electrodes to five\(^{13}\), corresponding to channels 3-7 in each subject’s map which spanned a -3 dB frequency range of 438-1063 Hz. The strategies for both those conditions employed the subject’s normal ACE filterbank and selected five maxima. The temporal envelope information coded by ACE5 was difficult to predict because the filterbank employs relatively narrow BPFs for apical channels. Consequently, the depth of modulation in each channel and the modulation phase across channels could vary depending on the distribution of harmonic frequency and phase relative to each BPF band (e.g., see differences in modulation depth across channels in

---

\(^{13}\) Although the five electrodes activated were fixed, slight variations in place coding did occur for these stimuli due to the interaction between harmonic frequencies and the channel band-pass filters. Measurement of the spatial centroid of the distributed electrical stimulation pattern across these stimuli revealed that the centroid only changed by up to approximately ±0.1 electrode place around the fifth most apical electrode.
Chapter 5: Assessment of the eTone Strategy

the third and forth panel of Fig 5.1). In addition, the coded modulation depth for modulation frequencies of 200, and in particular, 275 Hz was reduced because the -3 dB BPF bandwidths and envelope low-pass cut-off frequencies for these channels are 180 Hz. In contrast, because eTone applied envelope modulation based on the estimated F0, it was able to code F0s within its estimation range of 80 to 400 Hz. These two conditions represent typical modulation-rate coding that can be expected for ACE and eTone when presented with steady-state complex-harmonic or formant-like signals.

5.3.2 Results

Rate- and modulation rate-pitch DLs for reference F0s (or rates) of 125, 200, and 275 Hz are plotted in Fig. 5.2. Mean F0 DLs averaged across coding conditions for S1 and S5 (the musical/tonal-language experienced subjects) were 1 semitone or lower at F0s of 125 and 200 Hz and were less than 2 semitones at an F0 of 275 Hz. For S2, S3, and S6, mean F0 DLs averaged across coding conditions were between 1-2 semitones for F0s of 125 and 200 Hz but increased to greater than 3 semitones at an F0 of 275 Hz. For S4, mean F0 DLs were only less than 3 semitones at an F0 of 125 Hz.

Figure 5.2. Rate- and modulation rate-pitch DLs (in semitones) for experiment 1, plotted as a function of reference F0 rate for each coding condition examined (PulseTrain, ACE1, eTone1, ACE5, and eTone5) and for each subject, in panel (a), and the subject group average, in panel (b). For individual subject data, the error bars indicate standard deviation of the estimated thresholds. For group average data, error bars indicate 5% least significant difference (LSD) of means. Due to subject S6’s early withdraw from the study, DLs were not measured for ACE5 and eTone5 at all F0s, and for ACE1 at 275 Hz. For S3, DLs at 275 Hz for ACE5 and eTone5 were not measurable within the range 0.5 to 5 semitones. For four data points (denoted by *), estimated thresholds (DLs), derived from fitting of a psychometric curve, exceeded the largest F0 interval tested.
A three-way ANOVA was conducted on DLs to determine the effect of F0 and coding condition (PulseTrain, ACE1, eTone1, ACE5, eTone5). Normality and homogeneity of the data was observed. Significant effects of F0 ($F[2,10]=60.24; p<0.001$), coding condition ($F[4,18]=13.55; p<0.001$), and their interaction ($F[8,33]=6.71; p<0.001$) were found. Post-hoc comparisons showed that DLs for ACE5 at 275 Hz were significantly worse than those of all other coding conditions and F0s. In addition, within coding condition, DLs for ACE1 and eTone5 were significantly worse at 275 Hz than at 125 Hz. No other significant differences were found. Similar results to those described above were obtained when data from S6 (for whom the data set was incomplete) was removed from the analysis and when estimated DLs that exceeded the largest F0 interval tested (see Fig. 5.2 caption) were removed from the analysis.

5.3.3 Discussion

Subject-averaged unmodulated rate-pitch DLs (PulseTrain) presented at a single electrode were approximately 10 and 12% (with a range of approximately 3 to 22%) for rates of 125 and 200 Hz respectively. At 275 Hz, DLs increased to around 17% (with a range of approximately 6 to 30%). Those results are consistent with previous studies in which: (1) average rate-pitch DLs of 7.3% with a range of around 2 to 18% at a rate of 100 Hz were reported in a review by Moore and Carlyon (2005); and (2) rate-pitch DLs generally deteriorate as rate approaches 300 Hz (§2.3.2).

Similar DLs to those for the unmodulated pulse trains were obtained for amplitude-modulated stimuli processed by ACE1 (which employed a single wideband BPF) and eTone1 presented at a single electrode site. However, when signals were presented at five adjacent electrode places, only results for eTone5 were similar to those of unmodulated pulse trains. Results for ACE5 (which employed five narrowband BPFs) were significantly poorer at 275 Hz than those for all other coding conditions. The results for eTone5 demonstrate that the F0 modulation technique employed by eTone provides salient rate-pitch cues for F0s up to approximately 300 Hz, at least for the synthetic stimuli used in these tests. The comparatively poorer DLs for ACE5 show that is not the case for ACE. For F0s higher than 180 Hz, which exceed the BPF bandwidth and low-pass envelope cut-off frequency of apical channels in ACE, the depth of F0 modulation in the stimulus envelope of those channels decreases markedly. That
reduction in the modulation depth most likely accounts for the poorer DL observed at F0s of 275 Hz using ACE.

5.4 Exp. 2: Sung-vowel pitch ranking

5.4.1 Methods

Sung-vowel pitch ranking performance was measured using ACE and eTone. For each strategy, two numbers of maxima were evaluated (eight and ten) denoted as ACE8 and ACE10, or eTone8 and eTone10. The pitch ranking test employed the same 2I-2AFC procedure and instructions as described in experiment 1, in this case using pairs of sung-vowel tokens separated in F0 by three semitones. In each test block, eight repetitions of four F0 intervals were presented in random order. At least two blocks were administered for each subject and strategy in each session. The four F0 intervals were, 139-165, 165-196, 196-233, and 233-277 Hz, which corresponded to musical note intervals of C#3-E3, E3-G3, G3-A#3, and A#3-C#4 respectively. The vowel token was /a/ sung by a male choir singer and two recordings for each F0 were included in the test material. The duration of each token was 560 ms, which included a 30 ms linearly ramped onset and offset. An inter-stimulus interval of 500 ms was used. All tokens were equalised to a fixed average RMS level. Stimuli were presented in a sound-attenuated room via a loudspeaker placed 2.5 m from the subject at an azimuth of zero degrees. A maximum presentation level of 65 dB SPL was used, as measured at the subject’s microphone using A-weighting and fast response peak level metering. The stimuli were randomly presented at either 65 or 59 dB SPL to discourage the systematic use of loudness when ranking pitch.

The sung vowel token used in the present study was chosen because previous research (e.g., Vandali et al., 2005; Sucher and McDermott, 2007; §2.4.1.3) had shown that the vowel formant positions varied non-monotonically with F0, and therefore could introduce confounding pitch cues. An examination of stimulus spectrograms show ‘for example’ that F1 decreases slightly as F0 increases from C#3 to E3, but increases again for the interval E3 to G3. The test material was therefore considered difficult, but likely representative of everyday signals that contain unpredictable place/F0 combinations for which we would expect many CI users to perform poorly.
5.4.2 Results

Results of the sung-vowel pitch ranking tests for each strategy are plotted in Fig. 5.3. Discrimination results are shown for each F0 interval in terms of \(d'\) values calculated from bias-corrected percent correct scores (see methods in §3.3.1.1). For many cases in the individual data it seen that the G3-A#3 interval (196-233 Hz) was ranked incorrectly particularly when using ACE, for which a pitch-reversal was consistently reported. \(d'\) values for that interval were generally higher when using eTone although still not above threshold \((d' = 1)\) for all subjects. Of further note, for the poorer performing subjects, S2, S3, and S4, most F0 intervals were indiscriminable \((d' < 1)\) using ACE, whereas that was less often the case using eTone for the lowest two F0 intervals. The better performing subjects S1, S5, and to some extent S6, had difficulty in correctly ranking the first and third F0 intervals using ACE but ranked most F0 intervals correctly when using eTone.

Figure 5.3. Results for experiment 2 are plotted as \(d'\) versus F0 for pitch ranking of the sung-vowel stimuli for each subject, in panel (a), and the subject group average, in panel (b). For purposes of visual clarity amongst strategies, \(d'\) values at each F0 interval are joined by lines. For group average data, error bars indicate 5% least significant difference (LSD) of means. Group mean, bias-corrected, percent correct scores averaged across all F0 intervals are listed for each strategy in the legend.

Effects of strategy, number of maxima, F0 interval, and session were evaluated using ANOVAs. In all analyses, normality and equal variance of the data were observed unless otherwise specified. A two-way ANOVA with data averaged across strategy, number of maxima, and session, showed a highly significant effect of F0 Interval \((F[3,15] = 12.9; p < 0.001)\). Post-hoc Bonferroni tests for F0 interval revealed that \(d'\) for G3-A#3 (-0.37) was significantly worse than all other intervals (0.63, 1.20, and 1.34...
Chapter 5: Assessment of the eTone Strategy

for the first, second and forth F0 intervals respectively). Given the strong dependence on F0 interval, it was included as a main factor in all subsequent analyses.

The effect of session was examined for each strategy separately using three-way ANOVAs with data averaged across number of maxima. The results for S6 were not included in these analyses because she did not partake in the second sung-vowel test session. For both strategies, no significant effect of session was observed, although the trend more closely approached significance for eTone \(F[1,4]=5.2; \ p = 0.08\) than for ACE \(F[1,4]=0.33; \ p = 0.59\). Mean \(d'\) increased across sessions from 0.84 to 1.21 for eTone but dropped slightly from 0.28 to 0.23 for ACE. As expected, F0 interval was significant for ACE \(F[3,12]=19.0; \ p < 0.001\) with post-hoc test showing that \(d'\) for G3-A#3 was significantly worse than all other intervals. For eTone, F0 interval was weakly significant \(F[3,12]=4.2; \ p = 0.03\) with post-hoc tests showing that \(d'\) for G3-A#3 was only significantly worse than that for E3-G3. No other significant differences were observed. While the effect of session was not significant for the total subject group data, scores for the musically inexperienced subjects S2, S3, and S4 increased substantially across session using eTone. An ANOVA for strategy, F0 interval, and session for only those three subjects showed significant effects of strategy \(F[1,2]=66.0; \ p<0.001\), F0 interval \(F[3,6]=166.6; \ p<0.001\), and session \(F[1,2]=8.9; \ p=0.025\). A significant interaction between all three factors was observed \(F[3,6]=8.9; \ p=0.013\) with post-hoc tests showing that session was only significant for eTone for the two lowest F0 intervals.

The effect of number of maxima was examined separately for each strategy using three-way ANOVAs with data averaged across sessions. For ACE no significant effects of number of maxima was observed but as expected F0 interval was significant \(F[3,15]=16.1; \ p < 0.001\). Mean \(d'\) for eight and ten maxima were 0.45 and 0.25 respectively. For eTone no significant effects of number of maxima or F0 interval were observed. Mean \(d'\) for eight and ten maxima were 1.08 and 1.02 respectively.

Given that session and number of maxima were not found to be significant in the group analysis, the data were averaged across those factors to assess the effect of strategy. A three-way ANOVA for strategy and F0 interval showed significant effects of strategy \(F[1,5]=21.9; \ p = 0.005\), F0 interval \(F[3,15]=12.9; \ p < 0.001\), and their interaction \(F[3,15]=4.3; \ p = 0.02\). Mean \(d'\) for eTone was 1.05 compared to 0.35 for ACE. Similarly, the mean bias-corrected percent correct score for eTone was 73.3% compared to 57.2% for ACE. Chance performance for this test ranges from about 35 to 65% (according to a binomial distribution with N = 32). Post-hoc Bonferroni tests
Chapter 5: Assessment of the eTone Strategy

revealed that $d'$ scores for eTone were significantly higher than for ACE for the F0 intervals G3-A#3 ($p < 0.001$) and C#3-E3 ($p = 0.03$). Although normality was confirmed for this analysis, there was an indication that the assumption of equal variance was not justified ($p = 0.002$). A Friedman non-parametric ANOVA was therefore also used to assess the effect of strategy. Results confirmed the findings of the previous three-way ANOVA: For data averaged across all F0 intervals, a significant effect of strategy was observed ($p = 0.014$). Similarly, using the same approach for each F0 interval, a significant effect of strategy was found for G3-A#3 ($p = 0.014$) and C#3-E3 ($p = 0.014$), but not for the other F0 intervals.

5.4.3 Discussion

5.4.3.1 General outcomes

A significant effect of strategy was found in the sung vowel pitch ranking test. For group averaged results using eTone, all F0 intervals were consistently ($d' \geq 1$) ranked correctly except for G3-A#3 (196-233 Hz) for which $d' \approx 0.5$ indicating decreased salience of the pitch interval (although three of the six subjects, S1, S5, and S6, did rank this interval correctly). In contrast, for ACE only A#3-C#4 (233-277 Hz) was consistently ranked correctly ($d' \approx 1.5$). Performance for E3-G3 (165-196 Hz) $d' \approx 0.9$ and C#3-E3 (139-165 Hz) $d' \approx 0.2$ did not exceed the threshold criterion and G3-A#3 was consistently reversed ($d' < -1$) in pitch. These results demonstrate that eTone can provide substantial benefit to pitch perception when compared to clinical strategies such as ACE for the type of stimuli used in these tests. Furthermore, the laboratory analysis of eTone (see chapter 4, §4.3) combined with the rate/modulation-rate DLs from experiment 1, suggests that the benefit is most likely due to more salient and accurate coding of F0 information.

Subjects with musical/tonal-language experience (S1 and S5) obtained the lowest rate/modulation rate DLs in experiment 1 and for all coding conditions except ACE5 at 275 Hz their DLs were less than 2 semitones. Accordingly, those subjects also performed well in the sung vowel pitch ranking test in which stimuli were separated by 3 semitones. For the subjects with no musical experience, DLs for all coding conditions ranged from 2 to 3 semitones for rates of 125 and 200 Hz but increased to around 3 semitones or higher at 275 Hz. For those musically inexperienced listeners, results in
Chapter 5: Assessment of the eTone Strategy

the sung vowel pitch ranking test were generally reflective of their poor DLs, although some exceptions to this were observed, as discussed below.

Particularly for ACE, the results of the sung-vowel pitch ranking test can not be attributed solely to subjects’ F0 rate DLs or to the envelope low-pass filter in ACE. For instance, subject-averaged performance for the lowest F0 interval (C#3-E3) with ACE did not reach the $d’=1$ threshold criterion despite average F0 DLs of less than 3 semitones using ACE5 in experiment 1 at similar F0s. A likely explanation is that modulation rate information was poorly or incorrectly coded by ACE in the sung-vowel test, as was verified by examination of electrical stimulation patterns resulting from each stimulus using ACE, which often showed shallow F0 modulation in many channels. In addition, modulation in neighbouring channels was often substantially out of phase so that neurons responding to both channels were subjected to higher rates (e.g., $2 \times F0$) and reduced modulation depth. Performance for the highest F0 interval (A#3-C#4), which was consistently ranked correctly using ACE, is also not well predicted by the poor group average F0 DLs (> 4 semitones) when using ACE5 in experiment 1, and the attenuation of F0 modulation in ACE channels for F0s above the envelope low-pass filter cut-off frequency of 180 Hz. Examination of the spectral content of the A#3 token revealed the presence of weak F0/2 sub-harmonics which may have lowered the A#3 rate-pitch percept, thereby increasing the relative pitch difference between A#3-C#4 and decreasing (or reversing) that between G3-A#3.

The sung-vowel results also demonstrate the confounding influence of place of stimulation on listeners’ judgment of pitch, and support the conjecture that they used some weighted combination of rate and place of stimulation (McKay et al., 2000) to judge pitch, particularly for musically inexperienced listeners. This can be seen by examining the mean place of stimulation for each stimulus using a spatial-centroid model of the distributed electrical stimulation pattern (e.g., Laneau et al., 2006). For G3-A#3 the spatial centroid shifted by up to approximately 0.75 electrode places in the opposite direction to the change in F0. The conflicting place information in conjunction with the misleading F0/2 rate information for A#3 in this interval most likely accounts for the pitch reversal observed in results using ACE. For eTone, the conflicting place cues are equally present. However, because eTone only codes a single dominant F0, the weak F0/2 component in A#3 is eliminated. That factor combined with the deeper
Chapter 5: Assessment of the eTone Strategy

synchronous modulation provided by eTone, likely accounts for the improved $d'$ results compared to ACE for the G3-A#3 interval.

5.4.3.2 Effects of number of maxima

No significant effect of number of maxima was observed in the group data although average $d'$ values were slightly higher with eight maxima compared to ten. That outcome was unexpected as it was anticipated that inclusion of more maxima would reduce omission of spectral information and thereby produce fewer instances in which conflicting place cues to pitch were coded. The average change in place coding (measured via the spatial centroid) across F0 intervals was indeed smaller for selection of ten maxima compared to eight. However selection of eight maxima for these particular stimuli resulted in greater reinforcing, rather than conflicting, place-coding cues to pitch. Specifically, the higher scores for eight maxima compared to ten were most apparent in the results for E3-G3 where the change in the spatial-centroid was consistent with an increasing F0. Previous unpublished work by the authors had shown significantly poorer outcomes for sung-vowel pitch ranking tests with eight maxima compared to twelve. In those earlier tests, two of three subjects demonstrated a significant benefit for twelve maxima compared to eight when using the MEM strategy. Further work is needed to determine if better spectral sampling (e.g., by using more maxima or alternative channel selection algorithms) can reduce some of the confounding effects of place coding on pitch perception.

5.4.3.3 Effects of session

No effect of session was observed in the group data for either strategy, although for eTone the trend more closely approached significance than was the case for ACE. Improvement across sessions using eTone was apparent when only the data from musically inexperienced subjects (S2, S3, and S4) were analysed. For those subjects, no improvements across sessions were seen with ACE. For the experienced subjects (S1 and S5), ceiling effects in $d'$ for many of the F0 intervals were observed, even in the first session of testing. Given that both strategies received an equal/balanced exposure to the sung-vowel test, the significant effect of session using eTone for the musically inexperienced subjects cannot be attributed to task familiarity. It is more likely that those subjects learned to better utilise the enhanced rate-pitch information provided by the eTone strategy. However, whether this learning was facilitated by the intervening
rate-pitch DL tests of experiment 1 cannot be determined from this study due to lack of an adequate control. Further research is needed to determine whether directed rate-pitch training can be used to improve pitch perception in CI users.

5.4.3.4 Comparison to previous studies

The results with ACE in the present study are consistent with outcomes reported by Sucher and McDermott, (2007) for NH and CI listeners ranking the pitch of sung-vowels separated by one or six semitones. They observed that CI users of the SPEAK strategy (Skinner et al., 1991; Seligman and McDermott, 1995), obtained chance level performance in those tests. They also observed that musically experienced NH listeners performed significantly better than musically inexperienced NH listeners, with the latter performing at chance level for intervals of one semitone. In the present study when enhanced cues to F0 were provided by eTone, $d'$ scores for subjects with musical/tonal language experience (S1 and S5) were above chance level for all F0 intervals, whereas musically inexperienced listeners (S2, S3, S5, and S6) showed chance level performance for some of the F0 intervals. Although a direct comparison across studies is confounded by the different F0 intervals used in tests, average scores for S1 and S5 (92.2 and 93.0% respectively) for three-semitone intervals were as good as those obtained by the musically inexperienced NH listeners for six semitone intervals. Subjects S1 and S5 were further assessed with one semitone intervals and achieved scores of 75.0 and 70.1% respectively using eTone. Those scores were above the chance level threshold of 65.6% for that test and fell mid-way between performance reported by Sucher and McDermott for the musically experienced and inexperienced NH listeners. Whilst even higher scores might be achievable by CI users through more experience with eTone, it is unlikely that performance will reach that of musically experienced NH listeners given that those listeners have access to fine spatio-temporal information in resolved harmonics, whereas CI users can only make use of envelope-based rate-pitch information.

5.5 Exp. 3: Speech perception

5.5.1 Methods

Speech perception tests were conducted using ACE and eTone with ten maxima (ACE10 and eTone10). The tests consisted of open-set monosyllabic CNC words (based upon the
phoneme set of Peterson and Lehiste, 1962) presented in quiet at 60 dB SPL, and open-set CUNY sentences (Boothroyd et al., 1988) presented at 65 dB SPL in eight-talker noise (Auditec, St. Louis, catalogue No. C146-MT Multitalker). The noise level was chosen for each subject (in session three prior to running the test) so that a score of around 50% was obtained when using ACE. The tests were conducted in a sound-attenuated room using pre-recorded CD material played back via a loudspeaker placed 2.5 m from the subject at an azimuth of zero degrees. For each subject, two lists of CNC words (each containing 50 words) and two lists of CUNY sentences (each containing 102 words) were administered per strategy per session.

5.5.2 Results

Results of the word tests in quiet and sentence tests in noise averaged across sessions are plotted in Figs. 5.4 (a) and (b) respectively for each subject and for the subject group using the ACE10 and eTone10 strategies. In all tests, normality of the data and the assumption of homogeneity were satisfied. Two-way ANOVAs on the percent correct scores for the word test and sentence test were performed using strategy as the main factor. No significant effects of strategy were observed. One-way ANOVAs were also performed on data for each subject individually. Again no significant effects of strategy were observed.

![Figure 5.4](image)

Figure 5.4. Results of experiment 3 are plotted as percentage words correct for each subject and for the group subject average using ACE10 and eTone10. CNC word scores in quiet are shown in panel (a). Words correct for CUNY sentences in noise are shown in panel (b). The SNR used for each subject in the sentence tests are shown below the subject labels. The error bars for the individual subject and group-averaged scores show the 5% LSD for each analysis performed.
5.5.3 Discussion

The results of the speech perception tests in quiet and multi-taker noise revealed no significant differences between ACE and eTone. This was an encouraging outcome given that subjects’ only experience with eTone was in the laboratory. The lack of deterioration in scores, especially for the sentence tests conducted at SNRs ranging from +12 to +4 dB, demonstrated that enhanced temporal coding of F0 information in eTone could be applied without adversely affecting speech recognition outcomes. Furthermore, given that for speakers of tonal languages, perception of musical pitch may be correlated with lexical tone discrimination (e.g., Wang et al., 2011) the results of the present study suggest that eTone may offer benefits to perception of tonal languages. Further work is needed to examine recognition of lexical tones, which have smaller F0 intervals, shorter durations, and rapidly changing tone contours than the stimuli used in the present study.

5.6 Conclusion

The rate and modulation-rate pitch DL measures and the sung-vowel pitch ranking data showed that all CI users evaluated in this study were able to utilise rate information to judge pitch. Significant benefits were obtained in the sung-vowel test using eTone compared to the conventional processing employed by ACE. The stimuli used in that test were representative of a natural sung voice embodying slight variations in spectral-timbre (and hence place coding), which present a difficult problem for CI users, particularly those with little musical experience. The poorer performance obtained using ACE (even for the better performers) confirmed that such clinical strategies do not provide sufficiently salient F0 information. In contrast, enhancement of F0 information in the stimulus envelope using eTone showed significant improvement in listeners’ abilities to rank the pitch of these stimuli. However, performance with eTone remained poorer than that of NH listeners with comparable musical experience, as was expected given the present limitations in coding of fine spatio-temporal information in CIs. Because the eTone processing did not adversely affect speech recognition in quiet or in noise, the strategy may benefit CI users in real-world situations involving perception of voice intonation, speaker identification, lexical tone in tonal languages, and pitch intervals in monophonic musical sounds. Further work is needed to examine this.
Chapter 6: Development of a Pitch Training Program

6 Development of a Pitch Training Program

6.1 Introduction

It has been well established that cochlear implant recipients exhibit poorer pitch (§2.4.1), melody (§2.4.2), and lexical tone perception (§2.5) compared to normal-hearing listeners. This has been attributed mainly to an inability to code fine spatio-temporal information in electrical hearing, which in normal hearing provides the principle cues to pitch (§2.2.1). In addition, while changes in electrical stimulation rate can be used to elicit changes in pitch (§2.3.2; §2.3.2.2) the percept is relatively weak and more sluggish (§2.3.3.2) compared to that derived from fine spatio-temporal information in normal hearing and can be adversely affected by place or level of stimulation (§2.3.2.5; §2.3.2.6; §2.3.2.7).

The present research aimed to investigate whether we could improve pitch perception in CI hearing by providing specific training designed to help listeners associate F0 with pitch, and resonant frequency information with spectral timbre, so that each of those musical attributes are better perceived. Previous studies with CI recipients have shown that training can improve listeners’ ability to identify melody (§2.7.3.1; §2.7.3.2) and instruments (§2.7.3.3). For instance, Galvin and colleagues (2007) found that training using a melodic contour identification paradigm improved listeners’ ability on that task and also generalised to improved identification of familiar melodies. However, for those tests, which employed synthesised stimuli, place of stimulation was held relatively fixed and outcomes may differ for natural musical sounds in which changes in spectral timbre are common. Such changes in spectral timbre have been shown to influence both normal-hearing and cochlear implant listeners’ judgement of pitch (§2.4.3.3; §2.4.1.3; §2.7.1), particularly for listeners that are not musically trained which suggests that musical training/experience can improve pitch discrimination even when confounding effects of spectral timbre are present.

The results obtained when comparing the experimental eTone strategy to ACE in a group of six adult CI recipients (chapter 5) provide further support for the use of training. In that study, pitch ranking tests were conducted using acoustic stimuli processed through the ACE and eTone strategies (§5.4). The stimuli comprised variations of a natural sung vowel that varied slightly in formant frequencies across different F0 values, and therefore introduced some confounding place cues to pitch. The
Chapter 6: Development of a Pitch Training Program

tests were conducted before and after approximately five sessions of psychophysical tests in which subjects gained experience in using rate and amplitude modulation-rate to discriminate pitch (§5.3). Pitch ranking results for the eTone strategy were significantly better after the psychophysical tests than at the start of the study, but only for those subjects that had no prior musical experience. In contrast, no significant changes were observed for listeners that were musically experienced, or for any subject when using ACE. The improvement for the musically inexperienced listeners was attributed to listeners learning to better utilise the enhanced rate cues in eTone compared to ACE (§5.4.3.3).

The body of literature related to perceptual training paradigms is extremely vast (see Schmidt and Bjork, 1992 and Smith, 2010 for reviews). With relevance to the specific aims of the present research a number of features common to successful training techniques have been identified. One of the most important features was the provision of feedback (McCandliss et al., 2002; Smith, 2010). Another feature was the use of highly variable stimuli (§2.7.2.1), such as multiple talkers and varied phonetic content rather than a single talker and fixed phonemes. A third feature was the use of perceptual fading (§2.7.2.2), beginning with the most extreme exemplars of stimuli and progressing to more natural sounding exemplars as perception improves. Positive effects of training have also been shown for when primary cues alone are trained while holding secondary, or distracting auditory cues fixed, followed by the introduction of graded variations in secondary cues so that the stimuli evolve into more realistic sounds (§2.7.2.2). The training paradigm used in the present study embodied aspects of all those features to improve listeners’ ability to perceive pitch, particularly in music.

6.2 Pitch training paradigm

A computer-based training program in the form of a game known as aTune was developed. It required listeners to match pitch and/or spectral timbre to corresponding visual patterns. The stimuli differed in fundamental frequency and/or resonant frequency, which are the primary acoustic cues to pitch and spectral timbre respectively. The matching procedure involved listening to a sequence (Pattern) of one or more acoustically presented tones (Tokens) and then selecting a matching visual representation of the Pattern from a sequence of visual Tokens (referred to as a Block) on a computer screen. The matching procedure was repeated many times, each repeat constituting a
Chapter 6: Development of a Pitch Training Program

single Trial, over the duration of each training Run which lasted up to approximately five minutes.

A high-variability training paradigm (§2.7.2.1) was employed which comprised a large variety of synthesised complex-harmonic signals as well as recorded musical instrument and vocal stimuli. Further variability in the stimuli was introduced by varying secondary acoustic cues (§2.7.2.2; Iverson et al., 2005). Using an approach similar to perceptual fading and adaptive training (§2.7.2.2), training commenced with the simplest tasks (discrimination of an individual musical attribute using widely spaced stimulus intervals and holding other information fixed) and progressed to more difficult tasks using more natural sounds (discrimination of stimuli spaced by smaller intervals and discrimination of multiple attributes of musical signals). The size of the primary acoustic cue differences was adjusted according to each subject’s ability and progress through the training schedule. For instance for pitch training, the F0 interval was initially set to a size that the subject could clearly discriminate (as determined by discrimination tests conducted prior to the training). As the subject progressed through the training schedule the F0 interval was reduced. To promote generalisation of results to alternate tasks, the number of acoustic tones that differed within a training Run, and/or the number of tones presented sequentially within a matching Trial were also varied across the training schedule.

6.2.1 Training Stages and Levels

The training program comprised five Stages: (1a) pitch only training; and (1b) spectral timbre only training; (2a) combined pitch and spectral timbre training; (2b) instruments - pitch and timbre training; and (2c) instruments - pitch only training. The first two Stages employed stimuli in which discrimination of a single cue in isolation was trained whereas the next three Stages provided training of discrimination of multiple cues. Synthetic complex harmonic tone stimuli were employed in the first three Stages and recorded instrument stimuli in later Stages. Within each Stage, twelve different Levels were employed incorporating different stimuli, different parameter variations of the training task, and different visual layouts of the training screen.

6.2.2 Synthetic complex harmonic tone stimuli

The synthetic complex harmonic tones used in the first three Stages were generated using an additive synthesis technique from equal-amplitude sine-phase (i.e., zero initial
harmonics of F0 that were weighted in amplitude by a band-pass filter (see Fig. 6.1). Parameters of the band-pass filter that could be adjusted were resonant frequency, the -6dB bandwidth, and -26dB transition width. These parameters were all defined relative to F0 so that the spectral envelope was described in terms of harmonic number rather than absolute frequency.

![Diagram showing equal-amplitude sine-phase harmonics of F0 that were band-pass filtered by weighting the amplitude of harmonics according to parameters of a prototype band-pass filter shown in red.](image)

Figure 6.1. Equal-amplitude sine-phase harmonics of F0 (as shown at the top of the figure) were band-pass filtered by weighting the amplitude of harmonics according to parameters of a prototype band-pass filter shown in red. In this example, the band-pass filter has a resonant frequency of $F_0 \times 10$, a -6dB bandwidth of $F_0 \times 6$, and a transition width of $F_0 \times 2$.

### 6.2.2.1 $F_0$ variations

The permissible range of $F_0$s (i.e., the range of $F_0$s within the stimulus pool) was fixed between F2 (87.3 Hz) and B3 (247.0 Hz) which spanned a range of modulation rates that most CI subjects were expected to be able to discriminate (§2.3.3.1; §3.3.1; §5.3). For normal-hearing listeners and hearing aid users, that range was expanded to encompass
much of the bass and treble clefs of the western musical scale from F2 (87.3 Hz) to G5 (784 Hz). Within these ranges, the stimulus pool consisted of F0s corresponding to each note in the western musical scale and all subdivisions of the notes spaced by $1/16^{th}$ (0.0625) of a semitone.

6.2.2.2  Resonant frequency and bandwidth variations

The resonant frequency of the band-pass filter used to shape the spectral envelope of the synthetic tones was chosen from one of three possible ranges assigned to each Level: “Low” which spanned a range of 2-5 × the highest F0 of the tones to be presented within the Run; “Mid” (6-14 × highest F0); and “High” (10-31 × highest F0). Within each of these ranges, permissible resonant frequencies consisted of integer multiples of the highest F0 presented within the Run. One of two bandwidth profiles were also assigned to each Level: “Narrow” (bandwidth/transition width = 2/1, 3/1, 7/1 × highest F0 for the Low, Mid and High resonant frequency ranges respectively); and “Wide” (4/2, 6/2, 12/2 × highest F0 for the Low, Mid and High resonant frequency ranges respectively). See Appendix B for the resonant frequency ranges and bandwidth profiles assigned to each Level. For normal-hearing listeners that trained with a larger range of F0s, each of the three filter templates were modified so as to pass lower-order harmonics than those used for CI listeners (i.e., Low: 1-4 × highest F0; Mid: 4-8 × highest F0; and High: 8-14 × highest F0).

6.2.2.3  Temporal envelope variations

The temporal envelope applied to the synthetic tones varied across Levels. Four different envelopes were used: “Uniform” (fast 30 ms linear fade-in and 30 ms fade-out); “Swell-up” (linear fade-in over the duration of tone with a fast 30 ms fade-out); “Fade-out” (fast 30 ms linear fade-in with slow fade-out over the remainder of the tone); and “Transient-sustained” (Attack-Decay-Sustain-Release {ADSR} envelope with 15 ms attack time, followed by a decay to a level of 0.3 over 60 ms and ending with a 30 ms release). Schematic representations of each of these envelopes are shown in Fig. 6.2.
Figure 6.2. Four different temporal envelopes applied to the synthetic tones plotted as amplitude against time. The prototype for each envelope is shown by the solid blue lines. However because envelope were generated using an exponential envelope follower, the actual temporal envelopes are shown by the dashed black lines.

6.2.3 Musical instrument and sung-vowel stimuli

A pre-recorded database of musical instrument sounds (RWC: Real World Computing Partnership, Music Database - Musical Instrument Sound, Goto et al., 2003), was used to derive stimuli for Stages 2b and 2c of the training. Samples from three musical instruments: trombone; tenor saxophone; and piano; and from male and female rhythm and blues (R&B) vocalists, singing vowels: /a/, /i/, and /u/ were used. Specific details regarding the samples taken from the database are presented in Table 6.1.

All samples in the database were obtained from a DVD, as mono WAV files that were recorded with 16-bits of precision and a 44.1 kHz sampling rate. The samples taken from the database were edited using Sound Forge Pro 10 software so as to reduce their duration to 500 ms (which included 30 ms of linear fade-in and fade-out) and to normalise their RMS amplitude to -22 dB relative to full-scale (re-FS) for the three instrument sounds or to -12 dB re-FS for the sung vowels. The average F0 of each sample was determined using an autocorrelation based F0 estimator (using Praat version 5.3.03 software, www.praat.org) and checked against standard F0 frequencies of the western musical scale. If the estimated average F0 of any sample deviated by more than 0.125 semitones from the standard, it was adjusted accordingly using a pitch shift (Solo instruments 1) function of the Sound Forge Pro 10 software. Because samples produced by the tenor saxophone did not cover the full-range of F0s used in the training, three additional samples with F0s down to F2 (87.3 Hz) were produced using the same pitch shift function. The final edited set of samples comprised tones for each of the
Chapter 6: Development of a Pitch Training Program

Instruments and sung vowels spaced by one semitone on the western musical scale from F2 to B4. For intervals less than one semitone, a pitch-shifting routine was employed in the training program to shift F0 by steps of ± 0.5, 0.375, 0.25, or 0.125 semitones.

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Sample name</th>
<th>Articulation</th>
<th>Dynamics</th>
<th>F0 range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tenor Saxophone</td>
<td>271TSSTM</td>
<td>Staccato</td>
<td>Mezzo</td>
<td>G#2 - D#5 (103.8 - 622.2 Hz)</td>
</tr>
<tr>
<td>Trombone</td>
<td>221TBSTM</td>
<td>Staccato</td>
<td>Mezzo</td>
<td>E2 - A#4 (82.4 - 466.2 Hz)</td>
</tr>
<tr>
<td>Piano Forte</td>
<td>011PFSTM</td>
<td>Staccato</td>
<td>Mezzo</td>
<td>E2 - D#5 (82.4 - 622.2 Hz)</td>
</tr>
<tr>
<td>Male R&amp;B</td>
<td>501VMA1M</td>
<td>/a normal</td>
<td>Mezzo</td>
<td>F2 - G4 (87.3 - 392.0 Hz)</td>
</tr>
<tr>
<td></td>
<td>501VMI1M</td>
<td>/i normal</td>
<td>Mezzo</td>
<td>E2 - G4 (82.4 - 392.0 Hz)</td>
</tr>
<tr>
<td></td>
<td>501VMU1M</td>
<td>/u normal</td>
<td>Mezzo</td>
<td>E2 - G4 (82.4 - 392.0 Hz)</td>
</tr>
<tr>
<td>Female R&amp;B</td>
<td>501VFA1M</td>
<td>/a normal</td>
<td>Mezzo</td>
<td>E3 - E5 (164.8 - 659.3 Hz)</td>
</tr>
<tr>
<td></td>
<td>501VFI1M</td>
<td>/i normal</td>
<td>Mezzo</td>
<td>E3 - F5 (164.8 - 698.5 Hz)</td>
</tr>
<tr>
<td></td>
<td>501VFU1M</td>
<td>/u normal</td>
<td>Mezzo</td>
<td>E3 - F5 (164.8 - 698.5 Hz)</td>
</tr>
</tbody>
</table>

Table 6.1. Details of musical instrument and sung-vowel sounds taken from the RWC Music Database

6.2.4 Stimulus parameters varied across Stages

For the synthetic harmonic tones used in the first three Stages, four different temporal envelopes (§6.2.2.3) were used, but in early Levels, only the “Uniform” envelope was employed. For Stage 1a (pitch only), only F0 was varied within a training Run. Tone F0s were randomly chosen from those available (§6.2.2.1) to produce a fixed F0 interval. The BPF resonant frequency and bandwidth/transition width profile of stimuli was randomly chosen from those available (§6.2.2.2) for each Level, but was fixed across stimuli within a Run. In Stage 1b (spectral timbre only), only resonant frequency was varied within a training Run. The resonant frequencies of tones were randomly selected from those available to produce a fixed interval. Within any Run, tone F0 was randomly chosen from the allowable range and fixed amongst all stimuli. For Stage 2a (combined pitch and timbre), both F0 and resonant frequency were varied within a Run. As per Stages 1a and 1b, tone F0 and resonant frequency intervals were randomly chosen from those available for each to produce fixed intervals. For all three Stages, different ranges of resonant frequencies and BPF bandwidth profiles were used across Levels.

In Stages 2b and 2c, recorded musical instruments and vocal stimuli were used (§6.2.3). For these Stages, both F0 and instrument (or vowel) was varied within training Runs. Similar to Stage 1a, tone F0s were randomly chosen from those available to produce a fixed F0 interval. The selected instruments were fixed within each Level but
Chapter 6: Development of a Pitch Training Program

varied across Levels (see Appendix B). In Stage 2b, discrimination of both F0 and the instrument was required whereas in Stage 2c only F0 discrimination for stimuli that could differ in instrument was trained for.

For all Stages, the overall duration of tones used in the training was 500 ms and there was no interval between sequential presented tones. The intensity of all tones used within any training Run was RMS level balanced. In early Levels of each Stage, the intensity remained fixed. However, in later Levels up to 6dB of level roving was applied to each presented tone.

6.2.5 Discrimination procedure

The discrimination procedure involved matching of a sequential Pattern of acoustic tones to visual Patterns on a computer screen. The number of sequential tones within a Pattern varied from one to three across Levels. In addition, the number of primary cue variations (e.g., number of different F0s) within a Run varied from two to three across Levels. Variation of these two parameters provided control over the difficulty of each Level in terms of auditory, visual, and cognitive processing (e.g., Mousavi et al., 1995).

In earlier/easy Levels the trainee had a choice of one of two responses from presentation of a tone in a single interval. The task in that case can be classified as a single-interval, yes/no procedure, for which a percent correct score of approximately 69% corresponds to a d-prime (d’) value of 1 (Macmillan and Creelman, 1991) that is often used as a threshold criterion. D-prime is determined from the difference between the Z-transform of the hit and false-alarm rates, as per the equation 6.1.

\[ d' = Z(h) - Z(f) \]  

The response matrix below shows the results of an example training Run in which Tone 1 was presented 35 times, and Tone 2 65 times, and the trainee responded correctly 30 times when Tone 1 was presented and 50 times for Tone 2.

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Tone 1</td>
</tr>
<tr>
<td>Tone 1</td>
<td>30</td>
</tr>
<tr>
<td>Tone 2</td>
<td>15</td>
</tr>
</tbody>
</table>
For that example, the hit rate (i.e., correct proportion of responses for presentation of Tone 1) is 30/35 ($h = 0.86$) and the false alarm rate (i.e., incorrect proportion of responses for presentation of Tone 2) is 15/65 ($f = 0.23$) giving $d' = 1.82$.

In the next few Levels of each Stage, the number of sequentially presented acoustic tones was increased to two. Thus for each Trial there were four possible response Patterns (i.e., Tone: 1-1, 1-2, 2-1, and 2-2) making the task more difficult compared to the case in which only one tone was presented per Trial. If tones are treated independently, the task can be considered as two independent yes/no trials, and responses to both presented tones can be accumulated in one response matrix from which $d'$ can be estimated.

The same approach can be used for subsequent Levels in which the number of sequentially presented acoustic tones was increased to three. Later Levels of the training also increased the number of primary cue variations within a Run to three. Although the task is therefore absolute identification of three response alternatives, it can be treated equivalently by examining confusions between pairs of tone-levels and calculating corresponding $d'$ values in the same manner as described above. The number of primary cue variations and number of sequential tones within a Pattern for each Level of each Stage are listed in Appendix B.

### 6.2.6 Adaptation of tone intervals

In the first version of the training program (evaluated with normal-hearing listeners), the primary acoustic cue interval used in any given training Run was adapted according to the estimated $d'$ value (§6.2.5) obtained from the preceding Run. If the Level was completed successfully and $d'$ was greater than or equal to 2.0 (indicating that the trainee could discriminate the stimuli well), the interval size was reduced. For intervals greater than 6 semitones, the interval was reduced by 2 semitones. For intervals less than or equal to 6 semitones but greater than 1 semitone, the interval was reduced by 1 semitone. For intervals less than or equal to 1 semitone, it was halved. If $d'$ was less than 1.0, the interval size was increased in steps of 1 semitone for intervals $\geq 1$ semitone or by a factor of 2 for smaller intervals. Intervals were not increased beyond 8 times the subject’s measured DL. In addition, the adapted interval was applied to all subsequent Levels. However it was found that trainees had considerable difficulty progressing through Levels using that scheme. The adaptation scheme was thus modified in the final version of the program (which was evaluated with CI recipients) so all Levels
commenced with the same fixed interval size (i.e., four times the subjects’ DL determined prior to the training). Adaptation generally proceeded by reducing the interval according to the same step size rules described above if the Level was completed successfully and $d'$ was 2 or greater. If $d'$ was less than 2, the interval was usually not adjusted. Increases in the interval size were avoided unless a Level could not be completed after 20 attempts and $d' < 0.5$, in which case the interval was increased using the same step size rules described above. The largest F0 interval permitted was 12 semitones and the smallest 0.0625 (1/16th) semitones for synthetic tones or 0.125 semitones for recorded instrument stimuli. For synthetic tones, the largest resonant frequency interval permitted was 24 semitones and the smallest 0.125 (1/8th) semitones.

6.3 The pitch training game

In order to motivate listeners to maximise training and progression through the schedule, the training was embedded in a game play environment (aTune). To make the training more challenging, visual Tokens traversed along a predefined path defined by musical staffs on a computer screen. An example of one screen configuration used for pitch training is shown in Fig. 6.3. Selection of a visual Pattern involved positioning the screen cursor over the Tokens and pressing the left-button of the mouse. Selection of the correct visual Pattern eliminated that instance of the Pattern from the screen and thereby provided visual feedback regarding the correctness of the response. Selection of an incorrect visual Pattern inserted an additional Pattern to the screen, visually signalling that the response was incorrect. Upon selecting a visual Pattern (correct or otherwise) another Pattern was then presented acoustically which was randomly chosen from those available on the screen. If all visual Patterns were eliminated from the screen, or when a predefined time limit expired, a Block of new visual Tokens would enter the screen.
Chapter 6: Development of a Pitch Training Program

Figure 6.3. Screen capture for pitch training in Stage 1a. In this example, the tones are visually represented by quarter notes displayed on a musical staff. Visual Tokens representing higher F0 tones are positioned closer to the top of the staff than lower F0 tones. The visual sequences of Tokens enter from the top-right corner of the screen (see green arrow) and travel right-to-left along the top section of the path (see blue arrows which have been included in the figure to illustrate the progress of tones along the path), down the ramp, up the pipe (see dotted purple arrow), then change direction (left-to-right) and travel down the second ramp to drop off the end of the middle section of the path before finally travelling right-to-left to drop off the end of the bottom section of the path on to a drum (see red arrow). As an example, the visual Pattern of two Tokens that match those presented acoustically has been highlighted in blue to illustrate the correct Pattern that a trainee could select. Note that the visually correct Pattern can be located at more than one place on the screen (e.g., see second last Pattern of tokens at the top-right of the screen).

The object of the game was to eliminate visual Tokens before they reached the end of the path. Doing so for a period of five minutes and eliminating all remaining Tokens would result in successful completion of the Level. However, if any Token did reach the end of the path, the training Run would stop (recorded as a failed attempt) and that Level would need to be repeated.

6.3.1 Visual representation of tones

The acoustic tones were represented using different visual Tokens. For stages involving pitch perception only (i.e., Stages 1a and 2c), quarter note Tokens were used to depict tones with F0 (pitch-height) depicted by the vertical position on a musical staff. The example below shows three consecutive tones separated by 1 semitone. For small F0
intervals of less than 2 semitones, a “zoomed in” portion of the musical staff was used that included only 1 or 2 lines of the staff.

For Stages in which training targeted spectral timbre of synthetic harmonic tones (i.e., Stages 1b and 2a), timbre was depicted from dull to bright using the following three visual Tokens that varied in luminosity from dull to bright:

For Stage 2b in which training focused in identification of the instrument or sung vowel, timbre was depicted by an image of the musical instrument or by character mnemonics for each vowel:

6.3.2 Instructions and familiarisation

Prior to each training Run, a simple tutorial was provided in the form of a series of static images and text, describing the matching task. Information was provided about the number of different tones (i.e., primary cue variations) that would be heard within the Run and the number of Tokens within a sequential Pattern. A pictorial example of a Block of Tokens was shown along with an example of correct and incorrect selection of a Pattern from the Block. The tutorial also included both acoustic and visual presentations of each tone that would be used in the training Run. This allowed trainees to familiarise themselves with the primary cue variations used within the Run.

6.3.3 Mechanics of the game

6.3.3.1 Presentation of acoustic tone Pattern

The sequential Pattern of tones presented acoustically was always selected from one corresponding to a visual Pattern available on the screen. After each presentation of an acoustic tone Pattern, if the trainee made a selection (correct or otherwise), a new acoustic tone Pattern was presented approximately one second after the trainee’s selection. If however, the trainee did not make a selection within a given time interval,
the acoustic tone Pattern was repeated until a selection was made, at a rate that became progressively shorter as each Run progressed. The repetition interval began at two seconds at the commencement of each Run and decreased linearly to half a second by end of the Run.

6.3.3.2 Presentation and response time
Various parameters of the training game needed to be adjusted to provide suitable challenges that increased in difficulty with increasing Levels. Adjustment of those parameters needed to take into account the time taken to present sequential tone Patterns, expected response times, and the degree of difficulty desired for each Level.

The time required to complete presentation of each Pattern (PresentationTime) is given by the following equation:

\[ \text{PresentationTime} = (D + \text{ISI}) \times \text{NumSequentialTones} + L_1 \]  

where \( D \) is the duration of each tone (500 ms), ISI is the inter-tone interval (0 ms), \( \text{NumSequentialTones} \) is the number of Tokens presented acoustically within a sequential Pattern, and \( L_1 \) is a fixed delay (100 ms) representing average latency in the system.

The number of possible responses (i.e., number of combinations of tones within a Pattern: \( \text{NumCombinations} \)) is equal to the number of different tones (\( \text{NumTones} \)) raised to the power of the number of sequentially presented tones:

\[ \text{NumCombinations} = \text{NumTones}^{\text{NumSequentialTones}} \]  

Note that \( \text{NumTones} \) is the product of the number of different F0s (\( \text{NumF0s} \)) and the number of different resonant frequencies (or instruments/vowels) (\( \text{NumTimbres} \)) within the set of tones presented within a Run.

It was expected that response time (\( \text{ResponseTime} \)) would be dependent on the number of possible responses, processing time required for discrimination of the tones, and visual/cognitive processing (search) time. A linear equation modelling response time was used:

\[ \text{ResponseTime} = \text{R}_1 \times \text{NumCombinations} + \text{R}_2 \]
where $R_1$ reflects the processing time required to compare the acoustic tone Pattern with any visual Pattern from NumCombinations of possibilities and $R_2$ is a constant reflecting the time required to perceive/process the acoustic tone Pattern. These parameters were determined through trials with normal-hearing listeners and a few CI recipients. Values of $R_1 = 150$ ms, and $R_2 = 500$ ms, provided reasonable estimates of response times for the majority of those listeners and conditions tested. A detailed listing of expected response times is given for each Level of each Stage in Appendix B.

6.3.3.3 Appearance of visual Tokens

Each Token that entered the screen was randomly selected with replacement from those available within the set of Tokens corresponding to each acoustic tone used within a training Run. As such, all NumCombinations of visual Token Patterns were possible. Provided that the training time for the Run had not exceeded 5 minutes, a sequence of visual Patterns (i.e. a Block) entered the screen after a predetermined time had expired since the last addition, or when there were no remaining visual Patterns on the screen.

The number of visual Patterns within a Block (PatternsPerBlock) and the delay between additions of new Blocks (NewBlockDelay) were determined based on a function of presentation and expected response time. That function stipulated that the time interval between additions of each Block should be no less than the product of the time taken to respond to each Trial and the number of Patterns within a Block.

$$NewBlockDelay \geq \text{PatternsPerBlock} \times (\text{PresentationTime} + \text{ResponseTime}) \quad (6.5)$$

Appropriate (nominal) values for these parameters were determined to allow trainees to keep pace with the task and for control of Level difficulty and are detailed in the Appendix B for each Level/Stage. Within Runs, the time interval between additions of new Blocks decreased linearly from twice the nominal value at commencement of the Run, to the nominal value at the end of the Run. Accordingly, the number of Patterns within a Block increased linearly throughout the Run. To add variation within and between training Runs, randomisation of up to $\pm 50\%$ was added to these values.
6.3.3.4 Movement of visual Tokens

For each Level, visual Tokens travelled along a predefined path. Each Token in a newly added Block entered the screen one at a time to travel along the path. Many different path configurations were employed across Levels/Stages (see §6.3.8, also see Ramps path in Fig. 6.3).

The velocity of Token movement across the screen was determined according to the resolution of the screen and an estimate of trainees’ ability to track and select the moving Patterns using a computer mouse and screen cursor (a task related to hand-eye coordination). Through experimentation, a nominal Token velocity of around $\frac{1}{20}$th (0.047) of the screen width per second (or around 60 pixels/second for a screen width of 1280 pixels) was determined as a good compromise between challenge and achievability. A multiplicative scalar ($TokenVelocityScaler$) was used to vary that velocity across Levels so as to increase difficulty (see Appendix B). Throughout training Runs the Token velocity commenced at half the nominal value and increased linearly to the nominal value by the end. To improve visual appeal, the movement of Tokens was also accelerated or retarded according to the dynamics of the game. For instance, when Tokens first entered the path they were given a boosted velocity for a short period of time. Token velocity also changed slightly depending on whether the Tokens were moving up or down a ramp.

Tokens within each Block travelled as a connected group separated by a minimum distance equal to their visual width (approximately 30 pixels). If the leading Token of one Block came to within the minimum separation of the trailing Token of another Block, the two Blocks would be merged into one connected Block. Tokens within a Block could travel at different velocities but they would always de-accelerate or accelerate so as to maintain the minimum separation. For instance, when a Pattern was eliminated from the middle of a Block, the space created between Tokens was reduced to the minimum separation by decreasing the velocity of the leading Tokens for a short period of time. If Tokens were eliminated from the trailing portion of the Block, the leading Tokens would slow down by a greater amount than when Tokens were eliminated from leading portion of the Block. Changes in Token velocity accumulated when Tokens were eliminated consecutively and quickly and so Blocks of Tokens could be forced to travel at a negative velocity (i.e., backwards) if Tokens were eliminated quickly and consecutively.
Chapter 6: Development of a Pitch Training Program

several times. This provided an incentive to trainees to select *Token Patterns* as quickly and correctly as possible.

6.3.3.5 Selection of visual Pattern

Selection of a sequential *Pattern* of visual *Tokens* was carried out by positioning the screen cursor over the *Pattern* and pressing the left mouse button. For *Levels* in which there was only one tone presented per *Pattern*, the cursor needed to be positioned directly over the visual *Token* to be selected (within a tolerance of ±25 pixels). For *Levels* in which the number of sequentially presented tones was two, the cursor needed to be positioned between the two *Tokens* to be selected. For three sequentially presented tones, the cursor needed to be positioned over the middle *Token* in the *Pattern* to be selected. When positioning the cursor over *Tokens*, visual highlighting of those *Tokens* that would be selected was provided by changing their colour to blue. Upon selecting the *Token Pattern*, if the selection was correct the *Token Pattern* would be eliminated from the screen. If the selection was incorrect, a new *Token Pattern* would be inserted immediately after the first *Token* in the selected *Pattern*.

6.3.3.6 Completion of a Level

If the default 5 minute duration for a *Run* elapsed and no visual *Token* had reached the end of the path, new *Blocks* of *Tokens* would cease to enter the screen (although incorrect responses would continue to force the addition of new *Token Patterns*). If the trainee could subsequently eliminate the remaining *Tokens*, the *Run* would end and the *Level* would be flagged as successfully completed, otherwise it would be flagged as a failed attempt. The trainee could then move on the next *Level* if they desired, or to repeat the present *Level*, or any preceding *Level*.

6.3.4 Feedback

Visual and acoustic feedback was provided for all *Trials* to promote learning (§2.7.2.2). Visual feedback signalled a correct response by changing the colour of the selected visual *Pattern* to green for 500 ms before eliminating that *Pattern* from the screen. For incorrect responses, the selected visual *Pattern* turned red in colour for 250 ms after which a new visual *Pattern* was inserted after the first *Token* in the selected *Pattern* and it too was highlighted red for a brief time (250 ms). In early/easy *Levels*, the inserted *Pattern* depicted the correct response. In contrast, for later *Levels* the inserted sequence
was randomly assigned from the available set of *Patterns* (note, trainees were informed at the start of the *Run* as to whether the inserted *Pattern* signalled the correct response or not). Acoustic feedback in the form of a bright chime or a dull gong also followed correct or incorrect responses, respectively.

### 6.3.5 Hint to tone pattern

If the trainee did not respond within a given time after presentation of an acoustic tone *Pattern*, a correct visual *Pattern* was highlighted by changing its colour to aqua. That time delay varied between 6 seconds and 24 seconds depending on *Level*, expected response time, and progress through a *Run*. The hint delay at the commencement of a *Run* was shortest (typically 2 to 4 times the expected response time) and increased linearly with progress through a *Level* (typically to 4 to 8 times the expected response time).

### 6.3.6 Fast response rewards

To maximise the number of trials presented within a training *Run* and thereby increase the effective amount of training, a system rewarding quick consecutive correct responses was incorporated. The program employed a leaky integrator that increased in value by a given step size when a correct response was provided, or decreased by a given step size on an incorrect response. The integrator also started to decrease at a fixed rate after a given amount of time had elapsed post presentation without a response. If the integrator level exceeded a given threshold, a reward was issued and the integrator level was reset. The level of the integrator was displayed at the top of the screen using a horizontal bar so that trainees could observe the effects of their response latency and accuracy. Two parameters of the system were varied across *Levels* to control the ease of receiving a reward. The first was the threshold level for which a reward (*RewardThreshold*) was issued, and the second was the post presentation delay (*FastResponseTime*) after which the integrator started to decay. The *FastResponseTime* was adjusted so that it was always a little greater than the expected *ResponseTime* (§6.3.3.2). The *RewardThreshold* level was typically set so that 8 to 12 consecutive correct responses would trigger the reward.

The rewards available within each *Level* varied across *Levels*. Rewards could be won multiple times within a training *Run*. Rewards that were won within a *Run* did not carry over to the next attempt of the *Level*, or to any other *Level*. The rewards available were *Stop*, *Reverse*, *Blast*, *Zap*, *Hint*, *Slow*, *Hammer*, and *Cycle Target* (see Appendix
Chapter 6: Development of a Pitch Training Program

A). On receiving a reward, a corresponding button would appear at the top right of the screen. The reward could be activated at any time by positioning the cursor over the reward button and pressing the left-mouse button.

6.3.7 Bonus scoring mode and obstacles

In some Levels, anywhere from 5 to 20 “jewels” were distributed along the path. The jewels could be eliminated by timing a response such that the position of the selected Pattern coincided that of a jewel. If all jewels were eliminated within a Run, score increases from that time onwards would be doubled. Inclusion of jewels added another challenge to the game for those trainees who found the task easy and had fast response times.

In some Levels, 1 to 3 visual obstacles could further appear on the screen. These obstacles moved at variable speeds and along random trajectories around the screen, making the task more difficult by obscuring Tokens from time to time. Obstacles could be eliminated by selecting a correct Pattern when its position coincided with an obstacle.

6.3.8 Visual paths

Twelve different path configurations were used in the training program which provided varying degrees of difficulty and helped to promote continued interest in progression through the game (see Appendix C). Variation in difficulty was produced through changes in cognitive and visual loading provided across Levels, in addition to the variations in auditory loading described earlier (§6.2.5). This was introduced under the assumption that to complete the Level, auditory processing efficiency would need to increase so as to compensate for any increases in cognitive and visual loading. For the simplest path (known as Pipes), visual Tokens travelled right-to-left across the screen when moving with a positive velocity. Right-to-left motion required reading Tokens left to right, which is consistent with western musical notation and many written languages. Another path, (Back and Forth), allowed Tokens to also travel from left-to-right. When travelling left-to-right, the visual order of Tokens was reversed and consequently increased the visual/cognitive loading needed to perform the task. Other paths (Ramps, Holes, Up and Down, and Round and Round) included diagonal, up-down, down-up, left-right, and right-left motion, which increased the difficulty of tracking and selecting visual Patterns and created some confusion when diagonal paths crossed. Difficulty was also increased by introducing discontinuities in the path by transporting Tokens to
different places (Void Border, Random Voids, and Random Mix). The position of discontinuities was randomised between Runs so that additional cognitive processing was needed to determining the path on each Run. Other paths increased visual loading by introducing sections within the path in which visual Tokens could bounce back along the path (One Bounce, Two Bounce, and Round and Bounce) making tracking and selection of Patterns more difficult.

6.3.9 Scoring performance

At the end of each Run, performance was scored to provide trainees with feedback regarding progress relative to their own high scores for that Level and those of other trainees. The score was determined from the number of correct responses within a Run (NumCorrect) excluding responses for which only the correct Pattern(s) remained on the screen. The score was scaled by the number of acoustic tones presented sequentially in a Pattern and a bonus scoring scheme according to the following equation:

\[
Score = NumCorrect \times NumSequentialTones \times Bonuses \times DoubleScore
\]  

(6.6)

The Bonuses multiplier was determined from performance related to fast response time, use of small intervals, successful and fast completion of the Level, and a high d-prime score. The DoubleScore multiplier was determined from the ratio of correct responses obtained after entering the double score mode (DoubleScoreNumCorrect, see §6.3.7) and the total number of correct responses as follows:

\[
DoubleScore = 1.5 + DoubleScoreNumCorrect / NumCorrect
\]  

(6.7)

The overall score and bonus scores awarded were displayed after each training Run. Medals were awarded if the Level was completed successfully when the interval (Interval) between the primary acoustic cue(s) of the tones was close to, or smaller than the trainees measured threshold (DL). A bronze medal was awarded when the Interval ≤ 2 × DL, a silver medal when Interval ≤ DL, or a gold medal when Interval ≤ 0.5 × DL.

6.3.10 Logging of results

A detailed log for each attempt of any Level was maintained in a text file for each trainee to track his or her progress. The log included details such as the time, date, and attempt
number for the training Run, specific details about the stimuli used (e.g., F0s and resonant frequencies, etc), program parameters (e.g., NumTones, NumSequentialTones, etc) for the Run. In addition, for each Trial, the selected Pattern and response time were recorded, from which details about the number/percentage of responses correct, the average response time, and $d'$ for each primary cue were determined. The number of fast response rewards, the completion status (completed, failed, aborted), and total training time were also recorded.
Chapter 7: Assessment of Pitch Training Program

7 Assessment of Pitch Training Program

7.1 Overview

The potential benefits of the pitch training program to perception of pitch and spectral timbre was examined in a pilot study with normal-hearing adult listeners using a prototype version of the program. Some improvement to discrimination of F0-pitch but not spectral timbre was observed as a result of the training. However, the subjects did not train for the recommended duration and so the program was revised to encourage greater participation. The revised version was subsequently evaluated with a group of adult CI recipients. That study examined whether training involving discrimination of F0 information exclusively for judgement of pitch, and resonant frequency information exclusively for discrimination of spectral timbre, could improve subjects’ discrimination of F0 and reduce the known adverse effects of changes in spectral timbre (place of stimulation) on their judgement of F0-pitch.

7.2 Assessment in normal-hearing listeners

A group of six normal-hearing adults and six control subjects participated in the pilot study. All subjects were recruited from the University of Melbourne, School of Audiology and Speech Pathology, were of a similar age, and were paid for their involvement in the study. Allocation of subjects to each group (trainee or control) was approximately balanced for musical experience. Performance was evaluated through F0 and resonant frequency discrimination over time with each group. Tests were conducted at the start of the study (session 1) and approximately 4 weeks later (session 2). After the first session, the trainees were issued with the training program installed on a laptop computer which they took home for a period of approximately 3-4 weeks. They were asked to train for approximately 30 minutes per day or around 3 hrs per week to achieve a total training duration of between 9 to 12 hrs. The preliminary version of the training program comprised only the first three Stages (1a: pitch only; 1b: spectral timbre only; and 2a: combined pitch and spectral timbre) of the program described in chapter 6 (see §6.2.1). Trainees were expected to train in each of these Stages for approximately 1 week.
Chapter 7: Assessment of Pitch Training Program

7.2.1 Evaluation tests

The discrimination tests were conducted using a two-interval, two-alternative forced choice (2I-2AFC) procedure and a method of constant stimuli. The stimuli were synthetic complex harmonic tones consisting of equal amplitude sine-phase harmonics of F0 that were band-pass filtered. They were similar to those used in the first three Stages of the training program (see §6.2.2). For F0 discrimination, stimulus pairs were centred at a reference F0 of 150 Hz and were spaced by 6, 5, 4, 3, 2, 1, 1/2, 1/4, 1/8, and 1/16 semitones. The band-pass filter applied to the complex harmonic tones had a resonant frequency of 300 Hz, a -6dB bandwidth of 300 Hz, and a -26dB transition width of 600 Hz. The same tones were used for resonant frequency discrimination, although in this case, F0 was fixed at 150 Hz and the interval between the resonant frequencies of filters centred around 300 Hz was varied to produce intervals of 12, 10, 8, 6, 5, 4, 3, 2, 1, and 1/2 semitones. Prior to each test, the smallest interval was determined for which the subject could consistently rank pitch correctly (at least 7 out of 8 times) according to F0, or spectral timbre correctly (on a scale of dull to bright) according to resonant frequency. Five stimulus pairs were then formed, which included that interval and the four next consecutively smaller intervals. All tones were RMS level balanced and had durations of 500 ms. Tones within each stimulus pair were separated by an interval of 300 ms. The tones were presented at maximum level of 65 dB SPL and were roved in level by up to -6 dB across presentations to reduce any systematic effects of loudness on judgement of pitch or spectral timbre. No feedback was provided during testing. Within each test run, ten repetitions of each interval were presented in randomised order (e.g., half with the lowest F0 first and half with it second for F0 discrimination) providing a total of 50 trials per run. At least two runs were administered for each test. Response-bias corrected percent-correct scores (P_{c,max}) were used to determine discrimination thresholds for a performance criterion of P_{c,max} = 76% corresponding to \(d' = 1\) using the same methods described in section 3.3.1.1.

7.2.2 F0 discrimination results

F0 difference limens (expressed in semitones) measured in sessions 1 and 2 are plotted in the top row of Fig. 7.1 for each subject and the average of each subject group.
Figure 7.1. F0 DLs for trainees, controls and each group average are shown in the top row of the figure. Note the different scales used for DLs across groups and for the group average. Also note the break in the F0 DL axis for trainee data. For group averaged DLs, error bars plot 5% least significant difference of means for effect of session within group. The normalised improvement in F0 DLs for trainees, controls, and their group averages are plotted on a -log_2 scale in the bottom row of the figure. For group averaged normalised improvements, error bars plot 5% least significant difference of means and asterisk symbols indicate significant differences across group.

Group mean F0 DLs for trainees decreased from 0.81 to 0.38 semitones across sessions in contrast to 0.12 to 0.11 semitones for the control group. While mean F0 DLs differed across treatment groups no significant effect of group, session, or the interaction between group and session, was observed in an analysis of variance (ANOVA) for those factors with subject as a random (block) factor. In addition for each session, no significant difference between groups was found despite the poorer initial F0 DLs for the trainees compared to controls. However for the majority of trainees, a trend of improvement in F0 DLs was seen which was not apparent in the data for control subjects. The significance of this trend was not captured by the statistical analysis because improvements in performance are a compressive function of baseline performance (i.e., as baseline performance improves there is less scope for training to
provide further improvement). A more appropriate measure of performance compares the relative change in DLs across sessions with respect to the pre-training DLs. These data are plotted in the bottom row of Fig. 7.1 as normalised DL improvements derived from $-\log_2(DL_2/DL_1)$ where the DL subscript refers to session number. A negative log base-two transform was used so that successive halving of DLs produced equal positive steps in the performance measure\textsuperscript{14}. For instance, a reduction in DLs from 6 to 3 semitones across session 1 to 2 corresponds to an improvement (reduction) in the DL by a factor of 2 and a normalised improvement of 1. An ANOVA examining the normalised improvement in DLs across sessions was conducted in which subject group was the main analysis factor and subject a random (block) factor. A significant effect of group ($F[1,10] = 9.6; \ p = 0.011$) was observed in which the mean normalised DL improvement for trainees was 0.9 compared to 0.15 for controls. Those normalised DL improvements correspond to a mean decrease in F0 DLs across session by a factor of approximately 1.87 for trainees compared to 1.11 for controls.

### 7.2.3 Resonant frequency discrimination results

Resonant frequency DLs (expressed in semitones) measured in sessions 1 and 2 are plotted in the top row of Fig. 7.2 for each subject and each subject group average. Group mean resonant frequency DLs decreased from 2.5 to 1.1 semitones across sessions for trainees and from 3.4 to 1.8 semitones for controls. An ANOVA for the effects of group, session, and their interaction, and using subject as a random factor, revealed a significant effect of session only ($F[1,10] = 11.89; \ p = 0.006$) indicating that DLs improved across session for both groups. In addition, no significant difference between groups was found in an analysis of the normalised improvement in DLs across session (shown in the bottom row of Fig. 7.2). The mean normalised improvement in DLs were 0.98 and 0.87 (or mean decreases in DLs by factors of 1.97 and 1.83) for trainee and control groups respectively. These data suggest a strong effect of task learning which may have obscured any training effects.

\textsuperscript{14} In addition, applying the log transform helped to satisfy the requirement of equally variant data which is assumed when performing an ANOVA.
Chapter 7: Assessment of Pitch Training Program

### Figure 7.2

Resonant frequency DLs for trainees, controls and each group average are shown in the top row of the figure. For group averaged DLs, error bars plot 5% least significant difference of means for effect of session within group and asterisk symbols indicate significant differences across session. Normalised improvement in resonant frequency DLs for trainees, controls, and each group average are plotted on a \(-\log_2\) scale in the bottom row of the figure. For group averaged normalised improvements, error bars plot 5% least significant difference of means.

#### 7.2.4 Analysis of training logs

Information pertaining to a subject’s participation in the training (e.g., training duration and progress through training Levels and Stages) was extracted from training logs to determine whether there were any improvements that could be made to the training program. The average number of times that subjects trained over the 3-4 week training period was 6.5 and the average duration of training was 0.44 hours per occasion. Total training duration varied from 1.0 to 6.8 hrs ($\mu = 2.7$ hrs, $\sigma = 2.1$ hrs) across subjects which was well short of the suggested training duration of 9 to 12 hrs. The group average training duration for Stage 1a (pitch only) was 1.5 hrs which was approximately twice that for Stage 1b (spectral timbre only) and Stage 2a (combined pitch and spectral...
Chapter 7: Assessment of Pitch Training Program

timbre). These differences were also reflected in the average number of \textit{Levels} completed by subjects within each \textit{Stage}.

Most trainees, excluding TS6 who had the poorest F0 DLs amongst all subjects, completed each of the earlier \textit{Levels} of each \textit{Stage} on their first few attempts and subsequently moved on to the next \textit{Level} or next \textit{Stage}. However, as trainees progressed through \textit{Levels} within a given \textit{Stage}, the F0 or resonant frequency differences between training stimuli used were adapted to smaller intervals (see §6.2.6). As such, difficulty increased rapidly and more attempts were required to complete subsequent \textit{Levels}. For instance, F0 intervals (in \textit{Stage 1a}) commenced at a group average of 4 semitones in \textit{Level 1} and decreased to around 1.4 semitones by \textit{Level 6}. Conversely, the average number of attempts of \textit{Level 1} was 1.8 which increased to 6.8 by \textit{Level 6}. These outcomes suggested that to encourage a greater duration of training and progression through all \textit{Levels}, the program needed be revised so that F0 and resonant frequency differences between stimuli for each \textit{Level} commenced at a discriminal interval that was adapted independently for each \textit{Level}, see §6.2.6.

7.2.5 Discussion and conclusion

For the synthetic complex harmonic tones used in the present study, relative improvements in discrimination of F0 by the normal-hearing trainees were significantly greater (almost a twofold improvement) compared to control subjects. However, absolute F0 discrimination for the two groups was not significantly different. This may in part be due to the poorer overall initial performance of the trainees compared to the controls, despite both groups being matched in terms of previous musical experience. Resonant frequency discrimination for both groups improved substantially across evaluation sessions and no differences in relative improvements between groups were found, suggesting a strong effect of task learning across sessions. For future evaluations, it was thus recommended that sufficient task training and repeated tests be provided to subjects so that a plateau in performance is obtained in their initial evaluation session(s). The relatively short training duration completed by most of the trainees (compared to the recommended training duration) may have been insufficient to promote substantial improvements. The training program was thus modified to encourage a greater level of participation in future studies.
7.3 Assessment in cochlear implant recipients

The effect of the pitch training program in adult CI recipients using their everyday clinical device(s) was examined in two phases. In the first phase, subjects trained using synthetic complex harmonic tones in which a single cue to a musical attribute (i.e., F0 for pitch, or resonant frequency for spectral timbre) was varied. Results of tests conducted before and after the first phase were used to examine whether training with a single cue can improve a listener’s sensitivity to that cue in isolation (hypothesis 2a), and improve sensitivity to F0 in complex sounds in which multiple cues are varied (hypothesis 2b). In the second phase, subjects trained with synthetic and recorded musical stimuli in which multiple cues were varied. Results of tests conducted after the second phase were used to determine whether further improvements in a listener’s sensitivity to F0 in complex sounds, in which multiple cues are varied, can be obtained through training with complex sounds (hypothesis 2c).

The training program was modified from that used in the previous study with normal-hearing listeners. The main modifications included differences in: the manner and rate in which the “training intervals” were adapted (see §6.2.6); the parameter range of acoustic stimuli (e.g. upper F0 and frequency range of harmonics, see §6.2.2); the number of Stages used within the second training phase (see §6.2.1); and in the difficulty/complexity of game play (see §6.3 and Appendix B).

7.3.1 Subjects

Ten adult CI users, divided into two groups of five trainees (TS) and five controls (CS) participated in the study. The groups were balanced in terms of prior musical experience and modality of hearing. Due to the limited pool of subjects available, differences in gender, age, years of implantation, and implant/processor type, could not be balanced across groups (see Table 7.1). Three of the subjects also participated in the amplitude modulation psychophysics studies described in chapter 3. Those subjects were TS1, TS2, and CS3, who corresponded to subjects S6, S3, and S4 respectively for the study reported in chapter 3. Those subjects were thus more experienced at ranking pitch according to F0 rate than the other subjects.
Chapter 7: Assessment of Pitch Training Program

<table>
<thead>
<tr>
<th>Group</th>
<th>Sub</th>
<th>Gender</th>
<th>Age (yr)</th>
<th>Years implanted</th>
<th>Musical experience</th>
<th>Modality of hearing</th>
<th>Cochlear Implant</th>
<th>Clinical Processor</th>
<th>Clinical Strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trainee TS1</td>
<td></td>
<td>M</td>
<td>62</td>
<td>3</td>
<td>Music prior &amp; post CI</td>
<td>Bimodal</td>
<td>CI24RE (CA) CP810</td>
<td>ACE</td>
<td></td>
</tr>
<tr>
<td>Trainee TS2</td>
<td></td>
<td>F</td>
<td>66</td>
<td>2.5</td>
<td>Music prior to CI</td>
<td>Bimodal</td>
<td>CI24RE (CA) Freedom</td>
<td>ACE</td>
<td></td>
</tr>
<tr>
<td>Trainee TS3</td>
<td></td>
<td>F</td>
<td>51</td>
<td>2</td>
<td>None</td>
<td>CI only</td>
<td>CI512</td>
<td>CP810</td>
<td>ACE</td>
</tr>
<tr>
<td>Trainee TS4</td>
<td></td>
<td>M</td>
<td>50</td>
<td>8</td>
<td>None</td>
<td>CI only</td>
<td>CI24RE (CA) CP810</td>
<td>ACE</td>
<td></td>
</tr>
<tr>
<td>Trainee TS5</td>
<td></td>
<td>F</td>
<td>69</td>
<td>10</td>
<td>None</td>
<td>CI only</td>
<td>CI24M Esprit3G SPEAK</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control CS1</td>
<td></td>
<td>M</td>
<td>60</td>
<td>7</td>
<td>Music prior &amp; post CI</td>
<td>Bimodal</td>
<td>CI24RE (CA) CP810</td>
<td>ACE</td>
<td></td>
</tr>
<tr>
<td>Control CS2</td>
<td></td>
<td>M</td>
<td>83</td>
<td>8</td>
<td>Music prior to CI</td>
<td>CI only</td>
<td>CI24R (CS) Freedom</td>
<td>ACE</td>
<td></td>
</tr>
<tr>
<td>Control CS3</td>
<td></td>
<td>M</td>
<td>85</td>
<td>3</td>
<td>None</td>
<td>Bimodal</td>
<td>CI24RE (CA) CP810</td>
<td>ACE</td>
<td></td>
</tr>
<tr>
<td>Control CS4</td>
<td></td>
<td>M</td>
<td>65</td>
<td>12</td>
<td>None</td>
<td>CI only</td>
<td>CI24M Freedom SPEAK</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control CS5</td>
<td></td>
<td>M</td>
<td>79</td>
<td>11</td>
<td>None</td>
<td>CI only</td>
<td>CI24R (CS) Freedom</td>
<td>ACE</td>
<td></td>
</tr>
</tbody>
</table>

Table 7.1. Subject details for the five trainees and five controls. Trainees comprised 3 female and 2 males whereas all control subjects were male. Two subjects in each group had prior musical experience (highlighted in orange) and two were bimodal users of a CI and contralateral hearing aid device (highlighted in green). Subjects used different generations of the Nucleus 24 CI system and either the ACE or SPEAK clinical strategy. Average details for each subject group are highlighted in yellow at the bottom of the table.

7.3.2 Training schedule and evaluation protocol

All subjects were initially familiarised with the concept of pitch and spectral timbre in week 1 of the study (see the training/evaluation schedule summarised in Table 7.2). That familiarisation included tests involving ranking of pitch according to changes in F0, and ranking of spectral timbre according to changes in resonant frequency. These tests were repeated until a plateau in performance was observed so as to reduce the potential impact of task learning in subsequent tests on outcomes. In weeks 2 to 3, formal measurements of F0 and resonant frequency discrimination thresholds were conducted. The trainees were then issued with the training program installed on a laptop computer with audio output presented in the sound field by an amplified speaker (Behringer MS16). They were instructed on how to use the program and on the training schedule they should follow. They were asked to train using their everyday clinical device(s) (i.e., CI alone or CI+HA for bimodal subjects) for approximately 30 minutes per day or 60 minutes every two days.
Chapter 7: Assessment of Pitch Training Program

<table>
<thead>
<tr>
<th>Week</th>
<th>Phase</th>
<th>Evaluation Training</th>
<th>Control Group Tasks</th>
<th>Trainee Group Tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Familiarisation</td>
<td></td>
<td>Familiarise subjects with pitch, spectral timbre, and F0 &amp; resonant frequency discrimination tests</td>
<td>Familiarise subjects with pitch, spectral timbre, and F0 &amp; resonant frequency discrimination tests</td>
</tr>
<tr>
<td>2-3</td>
<td>1st evaluation session</td>
<td>1st evaluation session</td>
<td>Measure F0 &amp; resonant frequency DLs</td>
<td>Measure F0 &amp; resonant frequency DLs Issue training program and instruct on its usage</td>
</tr>
<tr>
<td>3-10</td>
<td>Phase 1 Training</td>
<td>N.A</td>
<td>Phase 1 self-administered training:</td>
<td>Phase 1 self-administered training:</td>
</tr>
<tr>
<td></td>
<td>Phase 1 Training</td>
<td>N.A</td>
<td>- Stage 1a (pitch only)</td>
<td>- Stage 1a (pitch only)</td>
</tr>
<tr>
<td></td>
<td>2nd evaluation session</td>
<td>2nd evaluation session</td>
<td>Remeasure F0 &amp; resonant frequency DLs</td>
<td>Remeasure F0 &amp; resonant frequency DLs</td>
</tr>
<tr>
<td>11-12</td>
<td>Phase 2 Training</td>
<td>N.A</td>
<td>Phase 2 self-administered training:</td>
<td>Phase 2 self-administered training:</td>
</tr>
<tr>
<td></td>
<td>Phase 2 Training</td>
<td>N.A</td>
<td>- Stage 2a (combined pitch/timbre)</td>
<td>- Stage 2a (combined pitch/timbre)</td>
</tr>
<tr>
<td></td>
<td>3rd evaluation session</td>
<td>3rd evaluation session</td>
<td>Remeasure F0 &amp; resonant frequency DLs</td>
<td>Remeasure F0 &amp; resonant frequency DLs</td>
</tr>
<tr>
<td>20-21</td>
<td>Extended Post Training</td>
<td>4th evaluation session</td>
<td>N.A</td>
<td>Remeasure resonant frequency and some F0 DLs</td>
</tr>
</tbody>
</table>

Table 7.2. Evaluation/training schedule. The study comprised a pre-training/familiarisation phase (aqua), training phase 1 (orange), training phase 2 (pink), and an extended post-training phase (purple). Evaluation sessions are highlighted in yellow and training periods in green. Listed in the table is the timeline for each of those phases in weeks, and the tasks to be carried out for each subject group.

The total duration of the training period, excluding evaluation sessions, was approximately 4 months (16 weeks). It was divided equally into two training phases. In the first phase (which addressed hypotheses 2a and 2b related to training using single cues), discrimination of F0 was initially trained in Stage 1a (pitch only) over a one month period (weeks 3-6), followed by discrimination of resonant frequency in Stage 1b (spectral timbre only) over a second one month period (weeks 7-10). For each Stage, trainees were first encouraged to complete Levels 1-6 and repeat those Levels many times so as to reduce the F0 or resonant frequency interval between stimuli (see §6.2.6) and improve their score (§6.3.9), before proceeding to Levels 7-12. After the second month of training (or approximately 2 months after the first evaluation tests for control subjects), all subjects returned to the laboratory for their second session of evaluation tests (weeks 11-12). Upon completion of those tests, the trainees embarked on phase 2 of the training (which addressed hypothesis 2c related to training with multiple cues) over another two month period (weeks 12-19). They were instructed to train for
approximately one month in Stage 2a (combined pitch and spectral timbre) using the same procedure outlined for earlier Stages and to spend the final month training in Stages 2b (instruments - pitch and timbre) and 2c (instruments - pitch only). Subjects TS1 and TS2 did not have access to Stage 2c and so they were only able to train with Stages 2a and 2b in phase 2. After the second phase of training (or after the second two-month period for controls), all subjects again returned for their third session of evaluation tests (weeks 20-21). Approximately three months after completion of the training (week 32), the trainees were further tested using a subset of tests to examine the effectiveness of the training on long-term outcomes. In addition, they were asked a series of questions related to their experiences when using the training program and the effect that it had on their appreciation of music and listening habits.

7.3.3 Initialisation of F0 and resonant frequency intervals for training

Before commencing the training, the initial interval between each primary acoustic cue to be used during training was established for each trainee. These intervals were derived from average F0 and resonant frequency discrimination thresholds measured in the first evaluation tests conducted prior to training. For each trainee, the initial intervals used for all Levels of the training program were established by increasing the measured thresholds (rounded to the nearest semitone integer) by a factor of approximately three to four so that the stimulus pairs would presumably be clearly discriminable. In addition, initial F0 intervals were restricted to a range of 4 to 8 semitones and initial resonant frequency intervals to 4 to 24 semitones. Table 7.3 lists the measured discrimination thresholds and initial intervals for each trainee.

<table>
<thead>
<tr>
<th>Trainee</th>
<th>F0 DL</th>
<th>Initial F0 Interval</th>
<th>Resonant Frequency DL</th>
<th>Initial Resonant Frequency Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>TS1</td>
<td>1.1</td>
<td>4</td>
<td>5.2</td>
<td>16</td>
</tr>
<tr>
<td>TS2</td>
<td>0.9</td>
<td>4</td>
<td>6.0</td>
<td>16</td>
</tr>
<tr>
<td>TS3</td>
<td>1.6</td>
<td>8</td>
<td>4.5</td>
<td>16</td>
</tr>
<tr>
<td>TS4</td>
<td>2.0</td>
<td>8</td>
<td>8.3</td>
<td>24</td>
</tr>
<tr>
<td>TS5</td>
<td>3.6</td>
<td>8</td>
<td>11.0</td>
<td>24</td>
</tr>
<tr>
<td>Group Av</td>
<td>1.8</td>
<td>5.6</td>
<td>7.0</td>
<td>19.2</td>
</tr>
</tbody>
</table>

Table 7.3. Measured DLs collected prior to training and initial primary acoustic cue intervals used in training. The F0 DLs are an average of DLs measured in tests 1 to 4 prior to training.
7.3.4 Evaluation tests

7.3.4.1 Stimuli and procedures

F0 discrimination was measured using six different tests. Stimuli used in tests 1 and 2 comprised synthetic complex harmonic tones in which only F0 varied. In test 3, a natural sung vowel was used for which formant frequencies remained relatively fixed among stimuli. Tests 4 and 5 comprised sung vowel stimuli with small variation in formant frequencies, and test 6 comprised two different sung vowels which introduced large differences in formant frequencies. Resonant frequency discrimination was measured in test 7 using synthetic complex harmonic tones in which F0 remained fixed and resonant frequency was varied. Specific details about the stimuli used in each test are listed in Table 7.4. Example stimulus output patterns (electrodegrams) for some of the stimuli used in these tests when processed by a 22 channel ACE strategy are shown in Fig. 7.3.

Tests 1-3 and test 7 served to examine the effect of training with a single cue (in phase 1) on subjects’ sensitivity to that cue. Tests 4-6 examined the effects of training (in phases 1 and 2) on F0 discrimination in complex sounds for which multiple cues varied. During the first phase of training, only synthetic complex harmonic tones were used. Therefore phase 1 evaluation tests that employed sung vowel stimuli (i.e., tests 3-6) provided a measure of generalisation of training effects to different stimuli. The F0 and resonant frequencies of stimuli used during evaluation tests were very rarely the same as those used during any phase of the training (which differed widely in F0 and resonant frequency across training Runs). It is thus unlikely that specific differences between stimuli learned during training, which were unrelated to the cue(s) being trained for, could be used by subjects to discriminate stimuli during tests.

Discrimination tests were conducted using the same 2I-2AFC procedure employed for normal-hearing listeners described in §7.2.1. Parameters of the synthetic stimuli (see §6.2.2) were adjusted for the cochlear implant recipients. For F0 discrimination, stimulus pairs were centred at a reference (nominal) F0 of 150 Hz and were spaced by 12, 10, 8, 6, 4, 2, 1, 1/2, 1/4 and 1/8 semitones. The BPF applied to the complex harmonic stimuli had a resonant frequency ($C_r$) of 300 Hz in test 1 and 1200 Hz in test 2 (see Table 7.4). The -6dB BPF bandwidth was 300 Hz and the -26dB transition width was 600 Hz. These tests served to measure F0 discrimination at two different spectral regions, the first involving mainly low-order (1-4) harmonics of F0 and the second
higher-order (4-12) harmonics. The same type of stimuli were used for resonant frequency discrimination in test 7, although in this case F0 was fixed at 150 Hz and the interval between the resonant frequency of BPFs centred around 300 Hz was varied to produce intervals of 16, 14, 12, 10, 8, 6, 4, 2, 1 and 1/2 semitones.

<table>
<thead>
<tr>
<th>Test</th>
<th>Discrimination</th>
<th>Stimuli</th>
<th>Varied parameters</th>
<th>Fixed parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>F0 (fixed place)</td>
<td>Synthetic complex harmonic tone</td>
<td>F0: 150 Hz nominal</td>
<td>BPF C_f = 300 Hz -6dB Bandwidth = 300 Hz</td>
</tr>
<tr>
<td>2</td>
<td>F0 (fixed place)</td>
<td>Synthetic complex harmonic tone</td>
<td>F0: 150 Hz nominal</td>
<td>BPF C_f = 1200 Hz -6dB Bandwidth = 300 Hz</td>
</tr>
<tr>
<td>3</td>
<td>F0 (relatively fixed place)</td>
<td>Sung vowel</td>
<td>F0: A2 to C3 F1: 0.23 Els F2: 0.25 Els</td>
<td>Male vowel /a/</td>
</tr>
<tr>
<td>4</td>
<td>F0 (small place variations)</td>
<td>Sung vowel</td>
<td>F0: A2 to C3 F1: 0.51 Els F2: 0.86 Els</td>
<td>Male vowel /i/</td>
</tr>
<tr>
<td>5</td>
<td>F0 (small place variations)</td>
<td>Sung vowel</td>
<td>F0: A3 to C4 F1: 0.48 Els F2: 0.47 Els</td>
<td>Female vowel /a/</td>
</tr>
<tr>
<td>6</td>
<td>F0 (large place variations)</td>
<td>Sung vowels</td>
<td>F0: A2 to C3 F1: 1.76 Els F2: 3.91 Els</td>
<td>Male vowels /a/ and /i/</td>
</tr>
<tr>
<td>7</td>
<td>Resonant frequency</td>
<td>Synthetic complex harmonic tone</td>
<td>BPF C_f: 300 Hz nominal</td>
<td>F0 = 150 Hz -6dB Bandwidth = 300 Hz</td>
</tr>
</tbody>
</table>

Table 7.4. Summary of stimuli used for discrimination tests 1-7. F0 discrimination was measured in tests 1-6 and resonant frequency discrimination in test 7. Synthetic complex harmonic tones were used in tests 1, 2 and 7 whereas sung vowel stimuli were used in tests 3-6. For F0 discrimination tests 1-3, place information (spectral timbre) remained relatively fixed amongst stimuli whereas for tests 4-5, place information varied slightly amongst stimuli, and for test 6, two different vowels were used within test which introduced large variations in place amongst stimuli. Stimulus parameters that were fixed within a test and those that were varied amongst stimuli are listed in the last two columns of the table respectively. For the male sung vowel tests, F0 varied between A2 to C3 (110.0 to 130.8 Hz) for all subjects except CS4 for whom the range was expanded to G#2 to G#3 (103.8 to 207.7 Hz). For the female sung vowel test, F0 varied between A3 to C4 (220.0 to 261.6 Hz) for all subjects except CS4 for whom the range was expanded to F#3 to G4 (185.0 to 392.0 Hz). For the sung vowel stimuli in tests 3-6, the maximum range of variation in mean place of electrical stimulation for the first (F1) and second (F2) formant frequency are listed in units of electrode place (Els) for the Nucleus 24 system (in which electrodes are separated by 0.75 mm).
Chapter 7: Assessment of Pitch Training Program

Figure 7.3. Stimulus output patterns (electrodograms) produced by a 22 channel ACE strategy to a few of the stimuli used in tests 2, 3, and 4. The top row of plots show the synthetic complex harmonic tones used in test 2. The middle row show the male sung vowel /a/ used in test 3 and the bottom row show the male sung vowel /i/ used in test 4. The left column of plots shows stimuli at a low F0 and the right at a higher F0. For all tests the F0 of the stimuli shown are separated by 2 semitones. Little variation in place of stimulation across F0s can be seen for those stimuli in tests 2 and 3. In contrast, small, but noticeable, place variations (particularly for F2) are observed for stimuli in test 4. Test 6 used both the male vowels /a/ and /i/ in randomised order. Large place differences for those stimuli are observed between the middle and bottom rows. Note, for these electrodograms, stimuli were presented directly to the speech processor's audio input and so F0 temporal information is arguably better represented in channels than would be the case for stimuli presented in the sound field.
Chapter 7: Assessment of Pitch Training Program

For tests involving F0 discrimination of sung-vowel stimuli (tests 3 to 6), vowels sung by a male, and a female rhythm and blues singer were used (see section 6.2.3). For most subjects, the F0 of the vowels used ranged from A2 to C3 for the male singer, and A3 to C4 for the female singer, although those ranges were expanded for subject CS4 (see Table 7.4 caption) whose performance on these tests was much poorer than that of all other subjects. Prior to each test, the smallest F0 interval for which the subject could consistently rank pitch correctly was determined. The two stimuli for that F0 interval, and two with equally spaced intervening F0s were selected for the test. For instance, if a subject could correctly rank the pitch interval between A2 and C3 (i.e., a 3 semitone interval), then A#2 and B2 were selected as intervening stimuli making a total set of four stimuli spaced by 1 semitone. Each of the six permutations of stimulus pairs for these four stimuli were used in the test (e.g., A2-A#2, A2-B2, A2-C3, A#2-B2, A#2-C3, and B2-C3). The smallest F0 interval between sung vowels stimuli used in tests was 1/4 of a semitone. For tests 3, 4, and 5, the same vowel (either /a/ or /i/) was presented within each trial, whereas for test 6, vowels /a/ and /i/ were both presented in balanced random order in each trial.

All synthetic tones and sung vowels were RMS level balanced and had durations of 500 ms. Stimuli were presented at a maximum level of 65 dB SPL (measured at the subject’s CI device) but were roved in level by up to -6 dB to reduce any systematic effects of loudness on judgement of pitch or spectral timbre. Pairs of stimuli in each trial were separated by an interval of 300 ms. Feedback was provided only during the familiarisation phase in week 1, and not during evaluation sessions. Within each test run, ten repetitions of each interval were presented in randomised order (half with the lowest F0, or dullest spectral timbre, presented first). Accordingly, that resulted in 50 trials per run for the synthetic tone tests 1, 2, and 7; 60 trials for the sung vowel tests 3, 4, and 5; and 120 trials for the mixed sung vowel test 6. For each test, at least two runs were administered. Response-bias corrected percent-correct scores were used to determine discrimination thresholds (for Pcmax = 76%) using the methods described in section 3.3.1.1.

7.3.4.2 Place variations amongst sung vowel stimuli

Differences in formant frequencies across stimuli used in sung-vowel tests 3-6 produced variations in place of stimulation that could influence listeners’ judgement of pitch (§2.3.2.5; §2.3.2.7; §5.4). The mean place of stimulation, or spectral centroid (Laneau,
Chapter 7: Assessment of Pitch Training Program

2004b), when processed through a 22-channel ACE strategy such as that used by the subjects in the present study, was measured for the first (F1) and second (F2) vowel formants of each stimulus. The third formant was much lower in intensity than F1 and F2 and varied little in place amongst stimuli, so it was ignored in this analysis. For the male vowel /a/ used in test 3, the maximum range of variation in mean place of stimulation was small for both F1 and F2 (i.e. less than approximately 0.25 electrode places, see Table 7.4 and the example electrogram plots in top row of Fig 7.3) compared to place differences that can be typically discriminated by CI users according to a review by Moore and Carlyon, (2005) (see §2.3.1.1). Those authors reported a median threshold of 1.2 mm, or 1.6 electrode places for users of the Nucleus 22 system, although thresholds as low as approximately 0.25 mm, or 0.33 electrode places, were reported for the better subjects and for Nucleus 24 CI subjects in a study by Laneau and Wouters (2004a). As shall be shown in section 7.3.5.2, resonant frequency (place) discrimination thresholds for the subjects in the present study were close to those of the better performing subjects in that review. Nevertheless, given that variations in place amongst stimuli in test 3 were smaller than the subject’s discrimination thresholds, it is unlikely that these variations influenced their judgement of pitch.

For tests 4-6, larger variations in mean place of stimulation were included. For the male sung vowel /i/ in test 4, mean place of stimulation varied by as much as approximately 0.8 electrode places because the singer altered articulation of the vowel /i/ slightly towards /e/ across F0s (see bottom row of Fig. 7.3). These variations were greater than the subject’s place discrimination thresholds and so may have influenced their judgement of rate-pitch. For the female sung vowel /a/ used in test 5, mean place varied by approximately 0.5 electrode places across F0s. However, because F0 for these stimuli was approximately one octave higher than for the male vowels, the weaker coding (i.e., average F0 modulation depth which was approximately 5% EDR for these stimuli) and poorer perception of F0 modulation rate at higher F0s (e.g., §2.3.2, §2.3.3, §2.6.3, and §5.35) means that variations in place may have influenced judgement of rate-pitch by a greater amount than for the male vowels. For test 6, the stimuli in each F0 interval tested consisted of two different vowels (/a/ and /i/) with even greater changes in formant frequencies across stimuli (see middle and bottom rows of Fig. 7.3). Outcomes of tests 4 and 5 were thus used to measure a listener’s F0 discrimination when accompanied by “small” variations in place (spectral timbre), whereas test 6 examined F0 discrimination for “large” changes in place.
Chapter 7: Assessment of Pitch Training Program

7.3.5 Evaluation test results

7.3.5.1 F0 Discrimination

Measured F0 difference limens (expressed in semitones) are plotted in Fig. 7.4. Individual subject data for trainees are shown in the left column for the six tests (shown in rows). Each plot shows measured performance at the three evaluation sessions. The middle column shows the same measures for the control subjects. Group averages are shown in the right column. Group mean DLs for each test and session are also listed in Table 7.5. For most tests, a trend of improvement in DLs between sessions 1 and 2 (i.e., after training phase 1) is seen for the trainees but not for the controls. Little difference in DLs between sessions 2 and 3 (i.e., as a result of training phase 2) was observed for either group of subjects. The effect of group (i.e., trainee or control) and session on DLs for each test was examined using an analysis of variance in which subject was a random (block) factor. For all F0 DL tests, no significant effects of main factors were observed but significant interactions between group and session were observed for test 1 ($F_{[2,16]} = 5.67; p = 0.014$), test 2 ($F_{[2,16]} = 3.88; p = 0.042$) and test 4 ($F_{[2,16]} = 4.27; p = 0.033$). In these cases, the 5% least significant difference of means showed that F0 DLs were significantly improved (i.e., lower) in sessions 2 and 3 compared to session 1 for trainees but not for the control group.

<table>
<thead>
<tr>
<th>Test</th>
<th>Group</th>
<th>Session</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Controls</td>
<td></td>
<td>4.2</td>
<td>2.8</td>
<td>2.2</td>
<td>2.5</td>
<td>2.1</td>
<td>2.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>4.1</td>
<td>3.2</td>
<td>2.3</td>
<td>2.7</td>
<td>2.0</td>
<td>3.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>4.7</td>
<td>3.1</td>
<td>2.0</td>
<td>2.3</td>
<td>2.0</td>
<td>2.4</td>
</tr>
<tr>
<td></td>
<td>Trainees</td>
<td></td>
<td>2.7</td>
<td>2.1</td>
<td>1.4</td>
<td>1.2</td>
<td>2.7</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>1.2</td>
<td>1.2</td>
<td>0.4</td>
<td>0.6</td>
<td>2.0</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>1.3</td>
<td>0.7</td>
<td>0.5</td>
<td>0.4</td>
<td>1.7</td>
<td>0.9</td>
</tr>
</tbody>
</table>

Table 7.5. Group mean F0 DLs in semitones for each test and session
Figure 7.4. F0 DLs for trainees, controls and each group average. For group averaged DLs, error bars plot 5% least significant difference of means for effect of session within group and asterisk symbols indicate significant differences across session relative to session 1. Subjects with unfilled symbols had bimodal (CI + HA) hearing.

To examine relative differences in DLs across sessions within subjects, normalised DL improvements were determined using the same procedure described in section 7.2.2. These data were derived from a logarithmic transform of the ratio of DLs across sessions as follows: $-\log_2(DL_n/DL_1)$ where the DL subscript refers to evaluation session number. Normalised DL improvements are plotted in Fig. 7.5 for each test, subjects and
the average of each subject group. Note that the normalised improvements for phase 1 were derived from session 2 DLs relative to session 1 whereas those for phase 2 were derived from session 3 DLs relative to session 2 and so that later only reflects improvements relative to those obtained after phase 1 training. Note also that an alternative analyse of relative improvements (using a logarithmic transform of DLs expressed as a percentage of the reference frequency) was performed in Vandali et al. (submitted). While that analyses technique differed from that presented here, outcomes of the analyses remained the same.

Separate ANOVA on the normalised DL improvement was conducted for each phase and test using group (trainees and controls) as the main factor and subject as a random (block) factor. For normalised F0 DLs in phase 1, a significant effect of group was observed for test 1 \((F[1,8] = 20.37; p = 0.002)\), test 2 \((F[1,8] = 7.26; p = 0.027)\), test 3 \((F[1,8] = 10.33; p = 0.012)\), and test 4 \((F[1,8] = 16.28; p = 0.004)\). Group mean improvements in normalised DLs for the trainees were 1.71, 0.91, 1.58, and 1.13 for tests 1 to 4 respectively. These normalised DL improvements correspond to mean decreases in F0 DLs by factors of 3.27, 1.88, 2.99, and 2.19 for tests 1 to 4 respectively. In contrast for the controls normalised DLs were effectively unchanged (0.05, -0.09, -0.10, and -0.31 for tests 1 to 4 respectively) across phase 1. While no significant effect of group was observed for tests 5 and 6, the mean improvement in normalised DLs for test 5 was 0.46 (or a factor of 1.38) for trainees versus 0.01 for controls, and for test 6 it was 0.55 (or a factor of 1.46) for trainees versus 0.01 for controls. These trends suggest that significant improvements for those two tests may also be exhibited in a larger subject pool. For normalised F0 DLs in phase 2, no significant effect of group was observed in any of the tests.

Further analyses were conducted in which data for tests with no place variations (1-3) were pooled, as were those for small place variations (4-5) (see Fig.7.6). For pooled data from tests 1-3, a significant effect of group was observed for phase 1 but not phase 2. For phase 1, the group mean normalised DL improvement for the trainees was 1.40 (or a factor of 2.64) compared to -0.05 for the controls \((F[1,8] = 78.40; p < 0.001)\). For pooled data from tests 4-5, a significant effect of group was also observed for phase 1 but not phase 2. For phase 1, the group mean normalised DL improvement for trainees was 0.80 (or a factor of 1.74) compared -0.15 for controls \((F[1,8] = 11.63; p = 0.009)\).
Figure 7.5. Normalised improvement in F0 DLs for trainees, controls, and each group average plotted on a -log base 2 scale. Data are plotted for the normalised improvement between session 2 and 1 (Phase 1) by the darker shaded bars and between session 3 and 2 (Phase 2) by the lighter shaded bars. For group averaged normalised improvements, error bars plot 5% least significant difference (LSD) of means for effect of group and asterisk symbols indicate significant differences across groups. Subjects that had bimodal (CI + HA) hearing are underlined in the axis legend.
Chapter 7: Assessment of Pitch Training Program

Figure 7.6. Normalised improvement in F0 DLs for test types in which place coding was relatively fixed (tests 1-3) or varied slightly amongst stimuli (test 3-5). For group averaged normalised improvements, error bars plot 5% least significant difference of means for effect of group and asterisk symbols indicate significant differences across groups.

7.3.5.2  Resonant frequency discrimination

Resonant frequency DLs (expressed in semitones) for test 7 collected in evaluation sessions 1, 2, and 3 are plotted in the Fig. 7.7 for each subject and the subject group. For all trainees, a trend of improvement in DLs between sessions 1 and 2 (i.e., after training phase 1) was seen. For most trainees DLs increased across sessions 2 and 3, although mean DLs for session 3 remained lower than those for session 1. For controls, little difference in group average DLs across sessions was observed, although a slight trend of improvement was seen for three of the five subjects. The same analysis model used for F0 DLs was applied to resonant frequency DLs. A significant effect of session ($F[2,16] = 3.75; p = 0.046$) and the interaction between group and session ($F[2,16] = 6.76; p = 0.007$) was observed. The 5% least significant difference of means showed a significant difference in DLs across all sessions for the trainee group (in which DLs for session 2 were lowest followed next by those for session 3) but no effect of session for the control group. The average group DLs for the three sessions were 6.98, 4.71 and 5.90 semitones respectively for the trainees, and 6.14, 6.47 and 6.08 semitones for the controls.
Figure 7.7. Resonant frequency DLs for test 7 plotted in semitones for each evaluation session. For group averaged DLs, error bars plot 5% least significant difference of means for effect of session within group and asterisk symbols indicate significant differences across session relative to session 1.

Normalised resonant frequency DL improvement are plotted for phase 1 and phase 2 training in Fig. 7.8. Separate analysis of variance were conducted on the normalised DL improvements in each phase using group as the main factor, and subject as a random (block) factor. A significant effect of group was observed for phase 1 \((F[1,8] = 12.60; p = 0.008)\) and phase 2 \((F[1,8] = 6.35; p = 0.036)\). For phase 1, the group mean normalised improvement for trainees was 0.58 (or a factor of 1.49) compared to -0.16 for controls. In contrast for phase 2, while the mean improvement for controls was sufficiently higher than that for trainees, the absolute level of improvement was small (0.095 for controls compared -0.26 for trainees).

Figure 7.8. Normalised improvement in resonant frequency DLs plotted on a –log 2 scale for phases 1 and 2. For group averaged normalised improvements, error bars plot 5% least significant difference of means for effect of group by accumulated phase and asterisk symbols indicate significant differences across groups.
Chapter 7: Assessment of Pitch Training Program

The average resonant frequency DL of subjects in the present study was approximately 6 semitones. When the stimuli used in this test are processed through a 22-channel ACE strategy, a change of approximately 6 semitones up, or down, from a resonant frequency of 300 Hz corresponds to centroid changes of +0.26, or -0.44, electrode respectively. These changes are comparable to the electrode place discrimination thresholds for better performing CI recipients reported by Laneau and Wouters (2004a) and Moore and Carlyon (2005) (§ 2.3.1.1). Furthermore, these data confirm that the resonant frequency DLs for the subjects in the present study were smaller than changes in place across stimuli in tests 4-6 but larger than those in test 3 and support the rational for subdivision of tests described in section 7.3.4.2.

7.3.5.3 Trainees’ results after an absence of training

To test whether the training provided any long-term improvement in trainees’ discrimination thresholds, some of the discrimination tests (2, 3, 6, and 7) were repeated approximately three months after completing the training schedule. Normalised DL improvements for the extended-post training session 4 relative to DLs obtained at the completion of training in session 3 and are plotted in Fig 7.9. Two-sided t-tests were conducted for each test to access the normalised improvements for the extended post-training phase. For all tests the mean normalised improvement was not significantly different from a mean of zero (test 2: \( t = 0.11, p = 0.914 \); test 3: \( t = -1.43, p = 0.226 \); test 6: \( t = -1.25, p = 2.81 \); and test 7: \( t = -0.29, p = 0.784 \)). T-tests were also conducted comparing the normalised improvement from session 4 relative to session 1. For that data, mean improvements were significant for test 2 (\( t = 4.04, p = 0.016 \)) and test 3 (\( t = 3.39, p = 0.028 \)) but not test 6 (\( t = 1.52, p = 0.202 \)) and test 7 (\( t = 2.37, p = 0.076 \)). Note however, that for all of these data, care should be taken in interpreting outcomes due to the low test power (\( N = 5 \)).
### F0 Discrimination Tests

#### Fixed spectral timbre
1. **Synthetic harmonic tones**
   - F0: Varied 150 Hz nominal
   - BPF C\(_f\) = 1200 Hz

#### Varied spectral timbre
2. **Sung Vowel: male /a/**
   - F0: Varied A2 to C3

3. **Sung Vowel: male /a/ & /i/**
   - F0: Varied A2 to C3

### Resonant Frequency Discrimination Test

5. **Synthetic harmonic tones**
   - F0: Fixed 150 Hz
   - BPF C\(_f\): Varied
   - 300 Hz nominal

---

**Figure 7.9.** Extended post-training normalised improvement in DLs for tests 2, 3, 6, and 7. Data are shown for the extended post-training session 4 relative to session 3. Error bars for the group average normalised improvement plot standard error of the means.

### 7.3.5.4 Bimodal versus CI alone listening

To test for differences in results for CI subjects who also used a hearing aid in their contralateral ear, separate ANOVAs were performed on data from trainees and controls, using modality of hearing (i.e., CI or CI+HA) and session as main factors, and subject as a random factor. For the trainees, a significant effect of hearing modality was observed for test 3 (\(F[1,3] = 30.59; p = 0.012\)) and test 5 (\(F[1,3] = 34.11; p = 0.010\)).
Mean DLs in both these tests were significantly lower for the two bimodal listeners compared to the three CI alone users. In addition, similar to the results seen in sections 7.3.5 and 7.3.5.2, a significant effect of session was observed in test 1 ($F[2,6] = 6.22; p = 0.034$), test 2 ($F[2,6] = 9.69; p = 0.013$), test 3 ($F[2,6] = 5.65; p = 0.042$), test 4 ($F[2,6] = 9.57; p = 0.014$), and test 7 ($F[2,6] = 9.67; p = 0.013$). No significant interaction between modality of hearing and session was observed in any of those tests. Thus, while mean DLs for the bimodal subjects were lower than those of CI-alone users in a few tests, the lack of any interactions between hearing modality and session across all tests suggests that changes in DLs across session were similar in trainees for both types of listeners. For control subjects, the only significant effect was the interaction between hearing modality and session for test 7 ($F[2,6] = 10.48; p = 0.011$). However, post-hoc analysis revealed that the mean resonant frequency DL for bimodal listeners was actually poorer in session 2 compared to session 1 in contrast to no effect of session for CI alone users. Those data are not indicative of an acoustic hearing advantage for bimodal compared to CI alone listeners. Similar conclusions regarding modality of hearing were obtained when the analysis was repeated for normalised DL improvements.

To test the extent to which bimodal listeners’ performance in discrimination tests prior the training was based on cues provide by their contralateral hearing aid ear, tests 3 and 4 were repeated using their CI device alone in evaluation session 1. DLs for these tests are plotted in Fig. 7.10. For each test, the effect of mode of hearing (i.e., bimodal or CI alone) was examined in an ANOVA using subject as a random factor. No significant effect of hearing mode was observed, although individual subject variability was evident in the data.

![Figure 7.10. Average DLs for tests 3 and 4 for bimodal and CI-alone listening in evaluation session 1.](image-url)
Chapter 7: Assessment of Pitch Training Program

That outcome was somewhat surprising given that perception of pitch in acoustic hearing (albeit impaired) is generally expected to be better than CI-only (§2.2.1, §2.4.2.2, §2.4.2.3). Aided acoustic audiograms for TS2, CS1 and CS3 demonstrated steeply sloping high frequency loss. TS1 had more usable acoustic hearing in the mid-frequency range, but very little hearing at low frequencies (see Table 3.2). For CS1, who is not included in that table, aided thresholds were approximately 30 dB HL below 500 Hz, 60 dB HL at 1 kHz, and greater than 75 dB HL at 2 kHz. Differences in outcomes across subjects are likely to be related to: their remaining low frequency acoustic hearing and hence their ability to use fine spatio-temporal information in low-order harmonics of F0 to derive pitch; and their ability to use temporal F0 information coded by their CI system as a cue to pitch. For instance TS1 (subject S6 in chapter 3), had demonstrated a good ability to use temporal cues to discriminate pitch in earlier experiments conducted in chapter 3.

7.3.6 Analysis of training logs

As was done in the previous study with normal-hearing listeners, information pertaining to subjects’ participation in the training was extracted from training logs to determine whether the changes made to the program were useful, and whether there were any additional improvements that could be made.

7.3.6.1 Training duration

The average number of occasions (days) on which subjects trained over the 16 week period was 55.4 (out of 112 days) and the average duration of training on those occasions was 0.52 hours. The total training time averaged across the five trainees was 28.8 hrs, with a range of 9 to 45 hrs and a standard deviation of 12.3 hrs. That average which was just over half the recommended training duration of 48 hrs, although more than sufficient to show an effect (e.g., Micheyl et al., 2006). The group average training durations for Stages 1a, 1b, and 2a were 12.6, 4.7, and 6.2 hrs respectively. The combined training duration for Stages 2b and 2c was 5.3 hrs. A detailed summary of training durations are shown in Fig. 7.11. When the difference in study duration between the present and previous study with normal-hearing listeners are taken into account, a substantial increase in subject participation is evident for the present study. However, a trend of decreasing participation with increasing Stages is present in both studies.
The correlation between training duration and evaluation test outcomes was examined using a linear regression model. Correlations with normalised improvement in DLs from each test and from the average of tests 1-3 (no place variations) and 4-5 (small place variations) were examined. Correlations were examined separately for data from phase 1 and phase 2. Table 7.6 lists those correlations that were significant or close to significant. Significant correlations were only found for data from phase 1. A significant correlation was observed between the normalised DL improvement for the average of test 4-5 in phase 1 (i.e., session 2 relative to 1) and the duration of training in phase 1. That correlation was mainly the result of the DL improvement in test 5 alone. In addition, that correlation was closely related to the duration of training in Stage 1a, which was almost significantly correlated with the DL improvement. These data suggest a relationship between training duration and F0 discrimination for some of the stimuli (namely those of test 5) that encompassed small variations in place. A significant correlation between the normalised improvement in resonant frequency DLs (test 7) and the duration of training in phase 1 was also observed. That correlation was mainly the result of the duration of training in Stage 1a and suggests improved discrimination of spectral timbre as a result of F0 discrimination training.
Table 7.6. Significant (p < 0.05) or near significant (0.05 < p < 0.08) correlations between training duration in Stages and evaluation test outcomes. Significant correlations are highlighted in yellow.

While some significant correlations were found, care should be taken interpreting those data due to the low sample size used in the analysis. In addition, because learning rates were not necessarily the same across subjects, training phases, and test measures, any significant correlations, or lack thereof, may well have been due to factors other than duration.

7.3.6.2 Progress through Levels

Table 7.7 lists details for each subject regarding the highest Level reached within a Stage, the total number of attempts at Levels within a Stage, the total number of successfully completed attempts across all Levels, and the average number of successfully completed attempts per Level for each Stage. As was observed when examining training duration across Stages, there was a trend of decreasing training effort with increasing Stage, particularly after the third Stage. For the first three Stages, all subjects completed Levels successfully more than once which is important because each time a Level was completed the interval between stimuli (i.e., F0 and/or resonant frequency) for that Level was reduced. On average in Stage 1a, subjects successfully completed 42 of 239 attempts at Levels, with an average of 4.5 completions per Level. These values decreased with increasing Stage number such that in Stage 2b, there were on average only 65 attempts and 22 completions, with an average of 2.1 completions per Level. It was encouraging however to observe that in prior Stages, subjects completed Levels enough times so that the intervals trained with approached their DLs measured in tests prior to the training.
Chapter 7: Assessment of Pitch Training Program

<table>
<thead>
<tr>
<th>Trainee</th>
<th>Stage</th>
<th>Highest Level</th>
<th>Levels Attempted</th>
<th>Levels Completed</th>
<th>Average Completions</th>
</tr>
</thead>
<tbody>
<tr>
<td>TS1</td>
<td>1a</td>
<td>12</td>
<td>332</td>
<td>60</td>
<td>5.0</td>
</tr>
<tr>
<td></td>
<td>1b</td>
<td>12</td>
<td>101</td>
<td>57</td>
<td>4.8</td>
</tr>
<tr>
<td></td>
<td>2a</td>
<td>12</td>
<td>191</td>
<td>49</td>
<td>4.1</td>
</tr>
<tr>
<td></td>
<td>2b</td>
<td>12</td>
<td>89</td>
<td>33</td>
<td>2.8</td>
</tr>
<tr>
<td></td>
<td>2c</td>
<td>12</td>
<td>273</td>
<td>30</td>
<td>2.5</td>
</tr>
<tr>
<td></td>
<td>1a</td>
<td>12</td>
<td>74</td>
<td>24</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td>1b</td>
<td>12</td>
<td>129</td>
<td>14</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td>2a</td>
<td>12</td>
<td>208</td>
<td>66</td>
<td>5.5</td>
</tr>
<tr>
<td></td>
<td>2b</td>
<td>12</td>
<td>226</td>
<td>53</td>
<td>4.4</td>
</tr>
<tr>
<td></td>
<td>2c</td>
<td>12</td>
<td>82</td>
<td>58</td>
<td>4.8</td>
</tr>
<tr>
<td></td>
<td>1a</td>
<td>12</td>
<td>101</td>
<td>35</td>
<td>2.9</td>
</tr>
<tr>
<td></td>
<td>1b</td>
<td>12</td>
<td>21</td>
<td>7</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td>2a</td>
<td>6</td>
<td>3</td>
<td>2</td>
<td>4.4</td>
</tr>
<tr>
<td></td>
<td>2b</td>
<td>6</td>
<td>68</td>
<td>34</td>
<td>5.7</td>
</tr>
<tr>
<td></td>
<td>2c</td>
<td>3</td>
<td>20</td>
<td>17</td>
<td>2.8</td>
</tr>
<tr>
<td></td>
<td>3a</td>
<td>5</td>
<td>8</td>
<td>20</td>
<td>4.0</td>
</tr>
<tr>
<td></td>
<td>3b</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>3c</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td>TS4</td>
<td>1a</td>
<td>7</td>
<td>298</td>
<td>33</td>
<td>4.7</td>
</tr>
<tr>
<td></td>
<td>1b</td>
<td>7</td>
<td>71</td>
<td>8</td>
<td>1.1</td>
</tr>
<tr>
<td></td>
<td>2a</td>
<td>5</td>
<td>121</td>
<td>7</td>
<td>1.4</td>
</tr>
<tr>
<td></td>
<td>2b</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>2c</td>
<td>3</td>
<td>126</td>
<td>2</td>
<td>0.7</td>
</tr>
<tr>
<td>TS5</td>
<td>1a</td>
<td>9.8</td>
<td>239.4</td>
<td>42.0</td>
<td>4.5</td>
</tr>
<tr>
<td></td>
<td>1b</td>
<td>9.8</td>
<td>69.6</td>
<td>32.8</td>
<td>3.1</td>
</tr>
<tr>
<td></td>
<td>2a</td>
<td>9.2</td>
<td>114.0</td>
<td>25.0</td>
<td>2.7</td>
</tr>
<tr>
<td></td>
<td>2b</td>
<td>6.8</td>
<td>65.2</td>
<td>21.6</td>
<td>2.1</td>
</tr>
<tr>
<td></td>
<td>2c</td>
<td>2.7</td>
<td>43.7</td>
<td>1.7</td>
<td>0.6</td>
</tr>
<tr>
<td>Group</td>
<td>Average</td>
<td>9.8</td>
<td>239.4</td>
<td>42.0</td>
<td>4.5</td>
</tr>
</tbody>
</table>

Table 7.7. The highest Level reached, the number of Levels attempted and completed, and the average number of completions per Level within each Stage.

7.3.6.3 *Stimulus discrimination and response time*

The ability of subjects’ to discriminate F0 and/or resonant frequency intervals used during training was measured via percentage of correct responses, dissimilarity index (\(d'\)), and response time. These values are listed for each Stage in Table 7.8 for each subject and the subject group average. Percent correct and \(d'\) values are shown as the average across all attempts as well as for completed attempts only within each Stage. Excluding TS5, \(d'\) values were nearly always greater than 1.0 indicating that most subjects were able to discriminate the stimuli correctly and average percent correct scores were on average around 70%. Results for Stages in which subjects had difficulty discriminating stimuli (i.e., \(d' < 1.0\)) are highlighted in orange. For those subjects and Stages, average percent correct scores were around 39%. For “completed” Levels only, average \(d'\) values were 2 or greater and percent correct scores were typically around 80% or higher. Because average \(d'\) values for attempted Levels were generally
(excluding TS5) greater than 1, but increased to 2 or greater for Levels that were completed, it can be concluded that cognitive loading arising from factors other than discrimination of the acoustic stimuli affected subjects’ abilities to complete Levels. As a consequence adaptation to stimulus intervals during training that were lower than subjects’ discrimination thresholds was also affected.

<table>
<thead>
<tr>
<th>Trainee</th>
<th>Stage</th>
<th>Average Percent Responses Correct (%)</th>
<th>Average d' Intervals</th>
<th>Average d' for Resonant Frequency Intervals</th>
<th>Response Time (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>For all Level Attempts</td>
<td>For Completed Levels</td>
<td>For all Level Attempts</td>
<td>For Completed Levels</td>
</tr>
<tr>
<td>TS1</td>
<td>1a</td>
<td>61.5 79.7 1.6 2.3</td>
<td>2.1 2.5</td>
<td>2202</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1b</td>
<td>72.5 82.9 2.7 3.3</td>
<td>3.0 3.4</td>
<td>2667</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2a</td>
<td>64.0 85.8 1.9 2.5</td>
<td>2.7 3.3</td>
<td>2615</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2b</td>
<td>70.1 83.0 1.9 2.5</td>
<td>2.7 3.3</td>
<td>2577</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2c</td>
<td>70.1 83.0 1.9 2.5</td>
<td>2.7 3.3</td>
<td>2577</td>
<td></td>
</tr>
<tr>
<td>TS2</td>
<td>1a</td>
<td>45.2 87.2 0.7 2.7</td>
<td>2.2 2.1</td>
<td>2411</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1b</td>
<td>76.2 90.4 2.1 2.5</td>
<td>2.2 2.2</td>
<td>2735</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2a</td>
<td>52.7 81.8 1.3 2.7</td>
<td>2.5 3.1</td>
<td>3575</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2b</td>
<td>60.8 84.7 2.0 3.0</td>
<td>2.1 3.1</td>
<td>5375</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2c</td>
<td>60.8 84.7 2.0 3.0</td>
<td>2.1 3.1</td>
<td>5375</td>
<td></td>
</tr>
<tr>
<td>TS3</td>
<td>1a</td>
<td>61.9 87.3 1.6 2.5</td>
<td>2.5 3.7</td>
<td>1228</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1b</td>
<td>87.4 94.6 3.1 3.7</td>
<td>3.7 3.1</td>
<td>962</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2a</td>
<td>69.7 86.2 2.0 2.7</td>
<td>2.5 3.1</td>
<td>1176</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2b</td>
<td>64.4 85.1 2.2 3.0</td>
<td>2.5 3.1</td>
<td>1513</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2c</td>
<td>84.7 91.5 2.8 3.3</td>
<td>2.5 3.1</td>
<td>1121</td>
<td></td>
</tr>
<tr>
<td>TS4</td>
<td>1a</td>
<td>74.6 86.3 2.0 2.6</td>
<td>2.5 3.0</td>
<td>1635</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1b</td>
<td>87.1 94.0 2.8 3.6</td>
<td>3.0 3.7</td>
<td>1268</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2a</td>
<td>78.1 94.0 2.8 3.6</td>
<td>3.0 3.7</td>
<td>1690</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2b</td>
<td>68.2 84.0 1.7 2.2</td>
<td>2.5 3.0</td>
<td>2245</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2c</td>
<td>70.5 90.2 1.9 2.6</td>
<td>2.5 3.0</td>
<td>1030</td>
<td></td>
</tr>
<tr>
<td>TS5</td>
<td>1a</td>
<td>46.4 78.1 0.9 2.1</td>
<td>1.3 2.9</td>
<td>548</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1b</td>
<td>54.5 88.1 0.4 2.1</td>
<td>1.3 2.9</td>
<td>1365</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2a</td>
<td>28.9 84.1 0.4 2.5</td>
<td>0.8 2.9</td>
<td>579</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2b</td>
<td>36.4 83.0 0.6 1.9</td>
<td>0.8 2.9</td>
<td>1642</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2c</td>
<td>38.9 61.4 0.4 0.9</td>
<td>0.8 2.9</td>
<td>643</td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>Average</td>
<td>57.9 83.7 1.4 2.4</td>
<td>2.3 3.2</td>
<td>1605</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1a</td>
<td>75.5 89.6 1.7 2.9</td>
<td>2.9 3.2</td>
<td>1729</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1b</td>
<td>75.5 89.6 1.7 2.9</td>
<td>2.9 3.2</td>
<td>1759</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2a</td>
<td>60.0 83.4 1.7 2.5</td>
<td>2.9 3.2</td>
<td>2670</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2b</td>
<td>58.7 86.4 1.7 2.3</td>
<td>2.9 3.2</td>
<td>931</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2c</td>
<td>64.7 81.0 1.7 2.3</td>
<td>2.9 3.2</td>
<td>931</td>
<td></td>
</tr>
</tbody>
</table>

Table 7.8. Average percent responses correct and d' values averaged across Levels within each Stage for each subject and the subject group average. Values are tabulated for the average of all Levels attempted, and for completed Levels only, within a Stage. Average response times (in ms) are listed for each Stage.

7.3.6.4 F0 and resonant frequency interval adaptation

For all Levels of each Stage, the F0 and resonant frequency intervals between stimuli were initialised to approximately four times the DLs measured in evaluation session 1 (section 7.3.3 and Table 7.4). As each Level was completed successfully, the intervals for that Level were reduced according to the rules described in section 6.2.6. The
average F0 and/or resonant frequency intervals for each Stage calculated for those Runs for which $d' > 1$, are listed in Table 7.9 along with the group averages. Also listed are the averages of the smallest F0 and/or resonant frequency intervals used within each Stage. These values were determined from an average of the smallest four intervals used in the last 33% of training Runs for a given Level in which $d'$ was greater than 1.0.

<table>
<thead>
<tr>
<th>Trainee</th>
<th>Stage</th>
<th>F0 Interval</th>
<th>Res Freq Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Average</td>
<td>Smallest</td>
</tr>
<tr>
<td>TS1</td>
<td>1a</td>
<td>4.2</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td>1b</td>
<td>4.3</td>
<td>2.75</td>
</tr>
<tr>
<td></td>
<td>2a</td>
<td>3.3</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>2b</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2c</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TS2</td>
<td>1a</td>
<td>4.2</td>
<td>3.0</td>
</tr>
<tr>
<td></td>
<td>1b</td>
<td>4.5</td>
<td>2.5</td>
</tr>
<tr>
<td></td>
<td>2a</td>
<td>2.8</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>2b</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2c</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TS3</td>
<td>1a</td>
<td>7.0</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td>1b</td>
<td>8.4</td>
<td>3.0</td>
</tr>
<tr>
<td></td>
<td>2a</td>
<td>9.2</td>
<td>7.0</td>
</tr>
<tr>
<td></td>
<td>2b</td>
<td>12.0</td>
<td>12.0</td>
</tr>
<tr>
<td></td>
<td>2c</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TS4</td>
<td>1a</td>
<td>5.9</td>
<td>1.75</td>
</tr>
<tr>
<td></td>
<td>1b</td>
<td>6.6</td>
<td>4.0</td>
</tr>
<tr>
<td></td>
<td>2a</td>
<td>8.3</td>
<td>8.0</td>
</tr>
<tr>
<td></td>
<td>2b</td>
<td>8.5</td>
<td>8.0</td>
</tr>
<tr>
<td></td>
<td>2c</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TS5</td>
<td>1a</td>
<td>10.42</td>
<td>5.0</td>
</tr>
<tr>
<td></td>
<td>1b</td>
<td>12.0</td>
<td>12.0</td>
</tr>
<tr>
<td></td>
<td>2a</td>
<td>12.0</td>
<td>12.0</td>
</tr>
<tr>
<td></td>
<td>2b</td>
<td>12.0</td>
<td>12.0</td>
</tr>
<tr>
<td></td>
<td>2c</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group Av</td>
<td>1a</td>
<td>6.3</td>
<td>2.7</td>
</tr>
<tr>
<td></td>
<td>1b</td>
<td>7.2</td>
<td>4.9</td>
</tr>
<tr>
<td></td>
<td>2a</td>
<td>7.1</td>
<td>5.6</td>
</tr>
<tr>
<td></td>
<td>2b</td>
<td>10.8</td>
<td>10.7</td>
</tr>
<tr>
<td></td>
<td>2c</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 7.9. Average and smallest F0 and resonant frequency intervals used during training.

Results for Stages in which a trainee’s performance improved over the duration of training are highlighted in orange if the smallest interval is at least 30% lower than the average interval, or in green if it is lower than the subject’s initial F0 or resonant frequency DL (see Table 7.4). For the latter cases, the subject was able to successfully complete a Level enough times to reduce the interval to their measured threshold or lower. Most subjects improved in one or more Stages of the training. Most improvement was observed in phase 1 (Stages 1a and 1b) although some improvement was also seen in phase 2 (Stages 2a and 2b).
Chapter 7: Assessment of Pitch Training Program

7.3.7 Post-training comments

At the completion of the study, trainees were asked about any problems they had experienced with the training program, whether they had found the experience enjoyable, and whether it had changed their appreciation of music and listening habits in anyway. A summary of their comments is provided in Table 7.10.

7.3.7.1 Criticism of program

All trainees experienced some initial confusion regarding how to perform the matching task. For instance, in the pitch training of Stage 1a, Level 1, pitch-height was visually represented by the vertical position of note symbols on a musical staff where higher positions corresponded to higher pitch. As is common in musical scores, the symbols for notes in the upper region of the staff are inverted compared to those in the lower region as is shown in the left panel of Fig 7.12. In doing this, the bottom of the symbol for the higher pitched tone is visually lower than that of the lower pitched tone (and vice versa for the top of the symbols). This introduced some confusion as to which symbol represented the higher or lower tone. After some feedback from trainees, the program was modified so as to always display note symbols using the same orientation regardless of position as shown in the right panel below.

![Figure 7.12. Position of note symbols on musical staff.](image)
# Chapter 7: Assessment of Pitch Training Program

## Criticism and Appraisal of Program

### Music appreciation and listening habits

<table>
<thead>
<tr>
<th>Trainee</th>
<th>Criticism and Appraisal of Program</th>
<th>Music appreciation and listening habits</th>
</tr>
</thead>
</table>
| TS1              | **Criticism:**
|                  | • Confused by inverted crotchets  |
|                  | **Appraisal:**
|                  | • Worthwhile                      |
|                  | • Enjoyable game                  |
|                  | • Improved pitch/reaction time    |
|                  | • Gained satisfaction from completing Levels & obtaining high scores/medals |
|                  | **Appreciation:**
|                  | • Can hear out single musical instruments and parts at live performances (e.g., Jazz) by concentrating harder. |
|                  | • No real change in musical activities since training (although now listens to some music DVDs). |
| TS2              | **Criticism:**
|                  | • Found some aspects of program confusing initially but worked them out with time. |
|                  | • Found combined F0 and resonant frequency (Stage 2a & 2b) difficult. |
|                  | **Appraisal:**
|                  | • Enjoyed the experience and challenge |
|                  | **Appreciation:**
|                  | • Pre-training music sounded awful using CI, but now music is more enjoyable with (CI+HA) but not as good as recollection of it with normal hearing. |
|                  | • Better concentration listening to music. |
|                  | • More confident to listen to live music |
|                  | • More confident to use CI+HA when listening to music (in the past would only use HA when listening to music) |
|                  | • Now encouraged to keep trying to get more pleasure from music. |
| TS3              | **Criticism:**
|                  | • Found some aspects of program confusing because e.g. couldn’t read music. |
|                  | • Found combined F0 and resonant frequency (Stage 2a, b and c) difficult. |
|                  | **Appraisal:**
|                  | • Enjoyed the game and experience |
|                  | **Appreciation:**
|                  | • Hear pitch better and can follow melody. |
|                  | • Can hear single instrument in pop songs. |
|                  | • Still can’t hear some things in music as remembered (e.g., off key or missing). |
|                  | • Now listens to new (unfamiliar) songs and enjoys them. |
|                  | • Has taken up learning the piano. |
|                  | • More confident to go to concerts. |
|                  | • Now listens to radio without audio cable. |
| TS4              | **Criticism:**
|                  | • Difficult to understand some concepts of the training game. |
|                  | • Frustrated - wanted to know more about what was happening and didn’t want to waste time working it out. |
|                  | • Found combined discrimination of F0 and resonant frequency very difficult. |
|                  | • More like a training task than a game |
|                  | **Appraisal:**
|                  | • Liked some challenges |
|                  | **Appreciation:**
|                  | • No real changes in appreciation of music |
|                  | • More confident to attend (and has attended more) musical performances since the training. |
| TS5              | **Criticism:**
|                  | • Frustrated by some Levels (particularly those with more than 2 tones) |
|                  | • Found spectral timbre discrimination (in Stages 1b and 2a) difficult |
|                  | • Found musical instrument identification very difficult |
|                  | **Appraisal:**
|                  | • Enjoyed the experience (“game” like except when frustrated where it was more like a “task”). |
|                  | • Mentally challenging and rewarding |
|                  | **Appreciation:**
|                  | • No real changes in appreciation of music |
|                  | • No real changes in listening activities |

Table 7.10. Subject comments regarding the training game and its effect on music appreciation/habits.
Subjects also reported some confusion when progressing from Level 1 of Stage 1a, which was an absolute pitch identification task, to Level 2 which required identification of both absolute pitch and direction of pitch change. Despite receiving instruction on how to perform these tasks, they still experienced some initial problems suggesting the need for a better tutorial system. Most subjects also commented that they found the combined discrimination of F0 and resonant frequency in Stages 2a and 2b difficult.

7.3.7.2 Appraisal of program

Most subjects, excluding TS4, enjoyed the training experience and commented that it was more like playing a game than a training task. Those subjects had little prior computer gaming experience whereas TS4 (being the youngest of the trainees and an experienced user of computers) was accustomed to more sophisticated games. His expectations of game play were higher than those of the other subjects which suggest that improvements to the game play and graphical aesthetics may be needed to provide better appeal to younger generations. TS4 also commented that he wanted to know more about the purpose of some aspects of the game/training design and the scoring/reward system.

7.3.7.3 Appreciation of music

Three of the five trainees commented that their appreciation of music had subjectively improved compared with their recollections prior to the training. For some, they felt that they could hear changes in pitch, follow melody, or could attend to individual instruments in music better. However, TS4 and TS5 indicated no change in their appreciation of music post-training. For TS4, this is consistent with his lack of positive appraisal for the program, his low training duration and poor progress through the training schedule (see §7.3.6.1).

7.3.7.4 Music listening habits

Most trainees indicated that their music listening habits had changed as a result of the training. Most commented that they were more confident about listening to music, particularly at live performances. One trainee (TS3) said that she now listens to new (unfamiliar) songs and enjoys them. That subject has also taken up playing of the piano since completing the training.
Chapter 7: Assessment of Pitch Training Program

7.3.8 Discussion and conclusions

For F0 discrimination tests 1-3, and resonant frequency discrimination test 7, discrimination of a single cue in the absence of other cue variations (excluding loudness) was examined. The significant improvements observed in these tests support the hypothesis (2a): that training with single cues can improve listeners’ sensitivity to those cues. The synthetic stimuli used in tests 1, 2, and 7 were similar to those used in the training, although during training a larger range of F0 and resonant frequencies, and different temporal envelopes were used. In contrast, the stimuli used in test 3-6 were sung vowels. The results from test 3 show that the effects of training generalised to discrimination improvements for more natural musical sounds in the absence of other cue variations.

F0 discrimination in the presence of small variations in resonant frequency was examined in tests 4-5. The outcomes partially support hypothesis (2b): that training involving discrimination of single cues will improve listeners’ sensitivity to F0 in complex sounds in which multiple cues vary. However, that hypothesis was not supported by results of test 6, in which much larger changes in resonant frequency were included amongst stimuli. Note that even for test 4 and 5, the resonant frequency place variations were above subjects’ discrimination thresholds (as confirmed in test 7 see §7.3.5.2). However in test 6 they were at least four times larger and may therefore exert a stronger effect on pitch. Furthermore, because F0 of stimuli in test 5 were an octave higher than those of test 4, and within the range in which temporal F0 information is poorly coded in CIs (§2.1, §5.3.3) and poorly perceived by implantees (§2.3, §5.35), the place variations in test 5 were likely to affect judgement of F0 pitch more than those in test 4, as was indeed observed in the results. In summary, the data show that training with F0 only variations can improve sensitivity to F0 in the presence of resonant frequency variations but to a diminishing degree as the relative effects of those place variations on judgement F0 rate-pitch increase.

The hypothesis that following single cue training, subsequent training with complex sounds in which multiple cues vary may further improve listeners’ ability to discriminate the pitch of complex sounds (hypothesis 2c) was tested in the second phase of the study. Outcomes of tests after the further training showed no improvement in DLs compared to tests conducted after single cue training. Because subjects only trained in phase 2 for approximately two-thirds of the duration that they did in phase 1, it is possible that greater improvements might be achieved with further training. That
conjecture is also supported by the significant correlations between training duration (in phase 1) and DL improvements (in test 5 and the average of tests 4-5). For tests 1-3 with relatively fixed place coding, no significant correlations between test outcomes and training duration were observed. It is likely that subjects’ performance in those tests saturated during the first phase of training and so irrespective of the type of training provided in phase 2, no further improvements were possible. For test 6 with large place coding variations, performance improvements were smaller and not significantly different across groups. For that test, training duration may have been insufficient to demonstrate substantial effects on DL, or the training itself may not have been effective.

In summary, while the results of tests 4-6 do not support hypothesis 2c, they do not explicitly reject it either.

Although significant improvements in resonant frequency discrimination were observed after the first phase of training, those improvements decreased after phase 2 such that they were no longer significantly different to those of the control subjects. In phase 1 almost a full month of training was prescribed for resonant frequency discrimination with no variation in F0, which was followed immediately by testing with comparable stimuli in evaluation session 2. In contrast, during the second phase of training, subjects were not trained on that task explicitly, and in the final Stages (2b and 2c) used a completely different set of stimuli (i.e., recorded instruments). It is thus possible that the decrease in resonant frequency discrimination from phase 1 to 2 can be attributed to lack of familiarity with the stimuli and/or with the task of ranking spectral timbre.

The long-term effects of the training on subjects’ improvements was examined by remeasuring F0 and resonant frequency DLs approximately three months after the training schedule was completed. No significant reductions in normalised DL improvements were observed in those tests for the extended post-training phase. However, mean values did decrease in some tests. For normalised resonant frequency DLs, decreased improvement was consistent with that which had already been observed from phase 1 to 2. Improvements in F0 discrimination remained more stable across phases but a trend of decreasing improvement was apparent in the extended post-training phase.

While results for the control subjects showed no significant change over time, mean DLs in some cases increased slightly across sessions compared to those for session 1. In the first weeks of the study, care was taken to ensure that all subjects were
well trained on the discrimination task so that any improvements in DLs observed in subsequent sessions could not be attributed to task learning. Tests were repeated until a plateau in performance was observed. In subsequent evaluation sessions, substantially less time was spent familiarising subjects with the task. This may in part account for the poorer DLs obtained by control subjects in sessions 2 and 3. In addition, control subjects may have become less interested with the task in later sessions. In contrast, trainees were more familiar with the task and probably remained better motivated to justify the amount of effort/time they had spent training. This raises the possibility that some of the benefits of training observed in the present study may have been due to factors related to subjects’ familiarity with stimuli and procedures, and their level of motivation.

Although only preliminary and for a small number of subjects it was observed that F0 and resonant frequency DLs were generally lower in subjects that had some residual hearing and used a contralateral hearing aid compared to those who used a CI alone. However, relative improvements in DLs across session for those subjects were no different to those for subjects with no residual hearing. Furthermore, when subjects with bimodal hearing were tested on some tests using their CI alone, little difference in DLs were observed compared to results when using both their CI and hearing aid. That outcome suggests that, at least for these subjects, their perception of F0 and resonant frequency was dominated by information provided by their CI.

7.3.9 Future research and recommendations

The further training in phase 2 with stimuli in which multiple cues were varied provided very little additional benefit after initial training with single cues in phase 1. However, the present study did not test whether initial training with simultaneous variation in multiple cues might provide similar benefits to those observed using single isolated cues. Nor did it test whether longer training with single cues might also provide further benefit compared to results of phase 1. Further research is needed to address these questions.

Most subjects commented that training involving multiple cue variations in the last three Stages was more difficult compared to earlier Stages. The increased difficulty most likely stemmed from having to discriminate pitch and spectral timbre simultaneously, rather than just pitch in the presence of varying spectral timbre. It may have in part been responsible for subjects’ lower training duration (§7.3.6.1), number of
Chapter 7: Assessment of Pitch Training Program

Level attempts (§7.3.6.2) and training improvement (§7.3.6.3) in later Stages. Assuming that training with multiple cues can be effective, the training procedures might benefit from improved balance of training difficulty across Stages. Possible approaches include: (1) limiting the task in Stage 2a and 2b to pitch discrimination in the presence of varying spectral timbre, and eliminating Stage 2c; (2) for Stages 2a and 2b, employ an adaptive training procedure that within each Run commences with target tone Patterns that differ in F0 but not resonant frequency, and gradually introduce Patterns in which both cues vary; and (3) for all Stages, employ a within training tutorial system that monitors responses and provides clearer instructions to address some of the comments raised by trainees, (§7.3.7) and to provide more informative feedback to promote faster learning.

Other possible changes to the program identified through analysis of training logs and subjective comments include: (1) reduction of Level difficulty by adjusting parameters that affect the velocity of visual Tokens, and the number of tone Patterns on the screen, (§6.3.3.3; §6.3.3.4). This would reduce cognitive loading arising from challenges other than discrimination of stimuli (§7.3.6.3) so that Levels are more easily completed and intervals between stimuli used during training approach the subject’s discrimination thresholds (§7.3.6.4); (2) revision of the scoring system to address subjects’ comments that their initial scores were very low compared to record high scores despite completing the Level with very few errors (§7.3.7.2). The present system heavily weights scores based on the inverses of the interval size used within a Run. In doing this, high scores could only be obtained once the interval is reduced to, or below, a subject’s discrimination threshold measured in prior to training; and (3) improvement to graphical aesthetics (e.g., varied background images, Token movement dynamics, and graphical effects) and game play (e.g., improved challenge and reward systems) to promote greater appeal to younger trainees.
Chapter 8: General Discussion and Conclusions

8 General Discussion and Conclusions

Motivation for this doctoral research was drawn from studies demonstrating poor pitch perception by cochlear implant recipients as compared to normal-hearing listeners. The primary reason for their poorer performance stems from the inability in existing CI technology to provide fine temporal and spectral information. Existing systems are capable of providing temporal and spectral envelope information from which pitch sensations can be derived. However, pitch salience, and height, derived from these mechanisms is relatively poor, and inaccurate, respectively, compared to that of normal hearing. The problem however does not end there. Coding of temporal F0 cues to pitch via electrical rate or modulation by existing clinical strategies is suboptimal. In addition, because electrical place codes information about spectral timbre rather than F0 pitch, judgement of rate-pitch can be affected when spectral cues vary independently of F0.

The present research addresses the poor coding and perception of temporal F0 information, and the potentially adverse effects of changes in place of stimulation on pitch. A two pronged approach to address these problems was adopted. The first aimed to improve presentation of F0 temporal information through the development of an experimental rate-pitch coding strategy known as eTone. The second sought to investigate whether training which directed listeners to attend exclusively to F0 information could improve pitch discrimination and reduce any adverse effects of place. These approaches aimed to optimise perception of pitch in individuals within the constraints imposed by electrical stimulation using present cochlear implant devices. Note that these optimisations primarily target perception of low F0's up to approximately middle-C on the western musical scale and pitch in monophonic signals.

8.1 Improved coding of rate-pitch using eTone

The experimental sound coding strategy (eTone) that was developed, sought to overcome weaknesses in existing clinical strategies related to salience and accuracy of elicited pitch. This was achieved through enhancement of temporal F0 information coded in the stimulus envelope. This enhancement included providing consistently deep F0 modulation in channels when they contained harmonics of F0, applying that modulation in-phase across all of those channels, applying that modulation in the form
Chapter 8: General Discussion and Conclusions

of a function characterised by a sharp-onset and rapid exponential decay, and deriving the modulation rate from an F0 estimator that was somewhat robust to effects of noise.

8.1.1 Pitch height

The depth and shape of the modulation function applied by eTone were chosen so that auditory neurons would be likely to fire in response to pulses at the onset of each F0 modulation cycle. Similar responses are expected to unmodulated F0-rate pulse trains for which strong phase locking to each pulse is observed (§2.3.2.1). Under the assumption that pitch-height is related to some weighted average of inter-spike intervals we might therefore expect both types of stimuli to produce similar pitch-heights. This is important because previous behavioural studies have shown that accurate identification of musical pitch intervals can be obtained by CI recipients on the basis of changes in electrical pulse rate, and so the same might be expected for changes in the modulation rate with eTone. In contrast, for sinusoidal or broad modulation functions similar to those produced by clinical strategies, a large temporal spread of neural responses within each modulation cycle is expected. For these modulators we might expect more substantial differences in pitch-height compared to those produced by F0-rate unmodulated and eTone-modulated high-rate pulse trains.

The above assumptions were tested in the present research through a series of psychophysical experiments that compared pitch elicited by low-rate unmodulated electrical pulse trains and amplitude-modulated high-rate stimuli. Results showed that for shallow modulation depths and low modulation rates, sinusoidal amplitude-modulated stimuli generally elicited a higher pitch than equally loud unmodulated pulse trains that were matched in pulse rate to the modulation rate (§3.4.1.2). However, as modulation depth was increased, the pitch-height of the SAM stimuli decreased towards that of the unmodulated low-rate pulse train. These data are consistent with the notion that rate-pitch is derived from a weighted average of inter-spike intervals that is increasingly dominated by responses to the modulation rate, rather than the pulse rate, as modulation depth increases (e.g., §2.3.3.4; §3.6.3).

The results from the above experiment did not however describe how the absolute pitch-heights of modulated stimuli compare to those of acoustic complex harmonic tones. This was addressed by comparing the two in a subset of the CI subjects who had some residual acoustic hearing in the contralateral ear. For those subjects, and the limited range of F0s examined (i.e. from 100 to 300 Hz), the results demonstrated a
Chapter 8: General Discussion and Conclusions

good correspondence between the pitch of the electrical pulse trains and the acoustic tones when pulse rate was equal to acoustic F0 (§3.5). Given that similar pitch-heights were also obtained between unmodulated and deeply-modulated electrical pulse trains matched in rate and modulation rate, respectively, suggests that deeply modulated pulse trains can elicit similar pitch-heights to those of complex tones in acoustic (albeit impaired) hearing.

In the above experiment which examined the effect of modulation depth, it was also observed that for low modulation rates, modulation depths of around 100% of subjects’ electrical dynamic ranges (EDRs) were often needed to produce close pitch matches between the SAM and unmodulated stimuli. Assuming that those outcomes are upheld when stimulating multiple electrodes in a sound coding strategy, these data confirm that the shallow coding of F0 modulation in clinical strategies is likely to elicit higher rate-pitch percepts than in normal hearing for low F0s, and that the relationship between F0 and pitch is likely to vary with changes in the depth and shape of modulation produced for different signals and acoustic environments (§3.6.2). In contrast, pitch-height comparisons between EDM stimuli (that comprise a sharp-onset and exponential decay modulation function similar to that employed in the eTone strategy) and unmodulated stimuli produced close pitch matches at shallower modulation depths of around 50% of subjects’ EDRs for low F0s. That result suggests that for harmonic sounds, the F0 modulation coded by eTone elicits pitch-heights much closer to those in normal hearing for low F0s than those produced by clinical strategies, at least in moderate to high SNRs for which the coded modulation depth remains relatively fixed.

The effect of modulation rate on relative pitch-height was also examined. Results demonstrated that for increases in modulation rate to 200-300 Hz, pitch differences between amplitude-modulated and unmodulated stimuli decreased (§3.4.1.2). That was to be expected if pitch percepts elicited by both unmodulated and modulated stimuli become saturated or less salient at high pulse or modulation rates, respectively, due to limitations in neural recruitment that arise from threshold and refractory effects (§2.3.2; §3.6.3). In another experiment, results showed that pitch differences between modulated and unmodulated stimuli were smaller for stimuli presented at a mid-loudness than at a high comfortable loudness level (§3.4.2.2). Those two outcomes suggest that the need to apply deep modulation in a sound coding strategy such as eTone is reduced at higher modulation rates and lower presentation levels. In addition, those two outcomes remain
Chapter 8: General Discussion and Conclusions

consistent with a model in which rate-pitch is derived from a weighted average of inter-spike intervals (e.g., §3.6.3).

A main aim of the present research was to determine ways by which adverse effects of place of stimulation on subjects’ judgments of pitch can be reduced. To examine the interaction between place and pitch, another psychophysical experiment was conducted to compare the relative pitch of two unmodulated pulse trains that differed in place of stimulation. For the majority of subjects tested, pitch was influenced by variations in place consistent with a tonotopic representation of pitch (§2.3.1; §2.3.2.5; §3.3.2). However, for a few subjects who were amongst those with the lowest pulse-rate discrimination thresholds, little effect of place on pitch was observed despite the fact that they could easily discriminate those changes in place-pitch under conditions of fixed rate. That latter outcome suggests that variations in the effect of place on pitch height across listeners arise from differences in the relative salience of rate and place cues to pitch, and that a weaker effect of place is expected when the rate cue is more salient (consistent with previous literature §2.3.2.5). The significance of this to present research is that improved salience of the rate cue in strategies such as eTone will reduce the potentially disruptive effects of place on pitch height. Furthermore, if the relative weighting of contributions from rate and place cues is subject to neural plasticity, then it may be possible to further reduce adverse effects of place through training.

8.1.2 Pitch salience

Previous research had shown that deep F0 modulation results in increased rate-pitch salience and improved rate-pitch discrimination thresholds (§2.3.3.1). The increased salience with increasing modulation depth presumably stems from narrowing of the temporal response distribution of neurons within each modulation cycle (§3.6.3). Accordingly, changes in modulation rate are more easily discriminated due to reduced temporal overlap between response distributions. Conversely, decreases in modulation depth reduce rate-pitch salience because of increased temporal overlap for the broader response distributions. Because the broader response distributions include more high-rate intervals, they can also produce the increase in rate-pitch height discussed earlier. It is known that clinical strategies often produce shallow and variable F0 modulation to different signals and acoustic environments (§3.6.2). Accordingly coding of rate-pitch is likely to be weak and inconsistent. Hence pitch-height is likely to be affected by variations in response distributions within each modulation cycle and variations in the
relative salience of rate and place cues. In contrast the application of deep/sharp F0 modulation using eTone is expected to increase rate-pitch salience and hence pitch height accuracy.

Rate-pitch salience is also affected by the spread of spatial excitation in electrical stimulation which introduces temporal integration of information presented at nearby electrodes, and a reduction and/or distortion of any F0 modulation that is presented out-of-phase at those sites (§2.3.2.8). Because the phase relationship between F0 modulation coded in channels of clinical strategies is not controlled, reduced and distorted coding of that modulation can result. To avoid that issue, eTone presented F0 modulation in-phase across channels that contained harmonics of the estimated F0.

Rate-pitch salience in clinical strategies can also be degraded by noise, which corrupts the temporal waveform in band-pass filter channels. This problem is partially alleviated in eTone by modulating channel envelope signals by a function devoid of noise. The function's modulation rate is derived from a F0 estimator that is robust to the effects of noise, at least for SNRs up to and including those typically encountered by many CI recipients. However, some noise can leak into channel signals if the harmonic-signal to noise ratio in those channels is too degraded (§4.2.3).

The effect of the F0 modulation enhancements introduced by eTone compared to the conventional processing produced by ACE was examined through a pitch ranking test with adult CI recipients (§5.4.2). In that test, two acoustic stimuli that differed in F0 were presented in random sequential order and subjects were asked to indicate which was higher in pitch. Outcomes demonstrated a significant improvement in ranking of pitch when using eTone compared to ACE, which was attributed to improved coding of F0 temporal information. The poorer performance using ACE (even by the better subjects) confirmed that such clinical strategies sub-optimally code F0 information, and that F0 modulation enhancements provided by eTone increase rate-pitch salience.

In the same study, psychophysical measures of rate- and modulation rate-pitch discrimination were also measured (§5.3). For low rates, DLs for most subjects in those tests were smaller than the F0 intervals used in the above pitch ranking test. Those data suggested that rate-pitch discrimination alone could not account for results of the pitch ranking test. Even when temporal processing in each strategy was taken into account, the results could only be explained by including effects of place. Further evidence to support that conclusion was seen in comparison of results with specific stimuli in the pitch ranking test. The stimuli consisted of natural sung vowel recordings with small
Chapter 8: General Discussion and Conclusions

variations in formant frequencies, and hence place of stimulation, that could adversely affect judgment of pitch. Evidence of that effect was observed in results for stimulus pairs that contained large place shifts in the opposite direction to the F0 cue. Using eTone, subjects that were musically experienced correctly ranked the pitch of all intervals above chance level performance, but failed to do so when using ACE for intervals that included the largest place shifts contrary to F0. Similarly, although the musically inexperienced subjects were unable to rank the pitch of all intervals correctly using either strategy, they obtained fewer errors when using eTone for those intervals with large place shifts (§5.4.3.1). If pitch is derived from a weighted combination of rate and place of stimulation and the adverse effects of place are reduced by increasing rate-pitch salience (in agreement with our earlier psychophysical findings in section 3.3.2 and data from other studies) the improved robustness to place variations indicates that eTone enhances rate-pitch salience compared to conventional processing in ACE.

8.1.3 Effects of eTone modulation on loudness

It was anticipated that loudness would be affected by the application of deep modulation in the stimulus envelope of eTone channels. Given that speech recognition by CI users can be sensitive to stimulus variations that affect coding of channel loudness (§2.6.3.2), correction for any loudness changes arising from application of deep modulation might be necessary. The effect of modulation depth on loudness was examined in chapter 3 through psychophysical experiments that determined the stimulation levels needed to balance the loudness of modulated and unmodulated pulse trains presented at a fixed place (§3.4.1.2). Outcomes showed that application of deep modulation while holding peak stimulation level fixed can reduce loudness by a significant portion of a subject’s electrical dynamic range. That was particularly evident at low modulation rates for the modulation function used by eTone. Assuming loudness is well coded and/or that subjects have become accustomed to the loudness coded by existing clinical strategies, that outcome indicates that some adjustment of levels may be needed in experimental strategies that code deep F0 modulation so as to maintain similar loudness. Previous analysis of voiced sounds processed by the ACE strategy showed that channel modulation was often sinusoidal-like, with an average effective modulation depth of approximately 10-20% of the EDR. For that modulation depth, the results of the loudness balancing experiments suggest that when processing signals using eTone, stimulation levels for an F0 of 100 Hz should be increased by approximately 14%
(range 4-24% across subjects) of the EDR to maintain loudness across strategies (assuming that loudness is integrated similarly across channels for both strategies). Outcomes also showed that loudness differences between deep and shallow modulated stimuli decrease with increasing modulation rate, which suggests that level compensation should decrease with increasing F0.

When fitting the eTone strategy in CI subjects, the adjustments necessary to balance the loudness of eTone and ACE for vowel sounds (§5.2.3) were well predicted by the results of the psychophysical experiments. Loudness differences between strategies were minimised by increasing the gain of F0 modulated signals in all channels by 1 to 6 dB across subjects, with an average increase of 3.8 dB. Those gain adjustments corresponded to increases in stimulation level ranging from 2.5 to 15% of subjects' EDRs with an average of around 10%, which was similar to the range and average of increases predicted by the psychophysical experiments. The similarity between predicted and measured adjustments provided some confidence that the adjustment procedure adopted to balance the loudness of strategies was appropriate. In addition, if a fixed gain of 3 dB is used as an alternative fitting procedure, the small gain errors of up to 3 dB predicted by these data may be acceptable.

**8.1.4 Speech perception using eTone**

When developing the eTone strategy, a great deal of attention was devoted towards ensuring that speech recognition was not compromised by the processing needed to improve coding of F0 information. In addition to matching loudness to ACE, the channel modulation algorithm was designed so that temporal envelopes applied to harmonic signals (e.g., voiced vowels) were deeply modulated according to F0, whereas inharmonic sounds (e.g. unvoiced consonants) retained the same envelopes as those normally produced by the ACE strategy. The performance of such an algorithm depends on how well harmonicity and F0 can be estimated and thus care was taken to develop a real-time F0 estimator that was accurate, had minimal lag, and was robust to the effects of competing noise. Laboratory analysis of the strategy established good accuracy of the F0 estimator (§4.3.2) and channel harmonic probability estimators (§4.3.3) for a variety of input signals with SNRs down to approximately 0 dB (which covered the range of SNRs commonly encountered by users of CIs). Those results ensured reliable estimates of F0 and modulation depth in proportion to harmonic probability.
Speech recognition performance with both ACE and eTone was examined with adult CI recipients. Outcomes in quiet and multi-talker noise revealed no significant differences between the two strategies (§5.5.2). The similarity in scores, especially for sentence tests conducted at SNRs as low as +4 dB, demonstrated that application of deep F0 modulation of the stimulus envelope to improve perception of pitch was practical and can be applied without adversely affecting speech recognition. However it was not clear whether the compensation for loudness differences between strategies was necessary. While not reported in chapter 5, speech recognition tests in quiet and noise were measured in three subjects without loudness compensation applied to eTone. For those few cases, average scores with eTone were lower than with ACE, suggesting that loudness compensation is important to retain intelligibility in CI users accustomed to using ACE. However, it remains to be determined whether that is the case for newly implanted subjects.

Another important indicator from the speech recognition results relates to the application of eTone to lexical tone recognition. If lexical tone discrimination by CI recipients is correlated with perception of musical pitch (§2.5.1.2), the results imply that eTone can improve lexical tone discrimination without adversely affecting recognition of segmental speech information in tonal languages. Some evidence of such benefits has already been observed in another study using a non-real time F0 coding strategy (§2.6.3.4), but further research in this area is required.

8.1.5 Effect of rate discrimination tasks on pitch perception

When evaluating performance of the eTone strategy, rate-pitch DLs at a single electrode place, and modulation rate DLs for stimuli processed through restricted channel conditions of the ACE and eTone strategies, were measured across a number of sessions (§5.3). Those tests were conducted in a psychophysical context without feedback and served mainly to determine a subject’s ability use different forms of temporal information to discriminate pitch. Results as a function of rate were fairly typical of those seen in similar studies showing an increase in DLs with increasing rate and the typical spread of performance across subjects. However, the data also showed that DLs may have improved across sessions, although that trend was not formally analysed in the study which only reported DLs averaged across sessions. Nevertheless those data suggested that either a subject’s ability to use temporal pitch information improved with experience, even when no feedback is provided, or that their ability to perform the task
itself had improved. To further explore that question, the sung-vowel pitch ranking test that was conducted at the start of the study, prior to the psychophysics, was repeated after the psychophysics test sessions had been completed. No significant difference across session using either ACE or eTone was observed. However for eTone, a trend of improvement across sessions was observed that more closely approached significance than was the case for ACE. That trend was due mainly to results of three of the six subjects who had no prior formal musical experience. For that group of three subjects, a significant effect of session was observed using eTone but not ACE. Given that both strategies received equal and order-balanced exposure to the sung-vowel test, the significant effect of session using eTone by the musically inexperienced subjects could not be attributed to task familiarity. It seems more likely that those subjects learnt to better utilise the enhanced rate-pitch information provided by the eTone strategy. Those results suggest that repeated exposure to tasks involving discrimination of rate and/or modulation rate in a psychophysical context (even without direct feedback) can improve musically inexperienced listeners’ discrimination of F0 in more complex harmonic sounds that vary also in place. That hypothesis was assessed using the training program developed in the present research (chapters 6 and 7) and is now discussed.

8.2 Training to improve pitch perception

In conducting this doctoral research, it became apparent that increasing rate-pitch salience through enhancement of F0 temporal cues is important, but of itself, not sufficient to overcome the adverse influence that place of stimulation can have on judgement of pitch. That is not surprising given that in normal hearing, auditory centres involved with perception of pitch have access to both fine spatio-temporal information and temporal envelope cues to F0. In CI hearing, centres that would normally process fine spatio-temporal information are not presented with fine-timing cues, and spectral information is coarse due to wide spread of excitation at the periphery. Temporal envelope cues are arguably better preserved in CI, but the cue is comparatively weak in normal hearing. When temporal envelope cues are weak, it is therefore expected that the coarse spectral cues will exert a strong influence on pitch.

It is possible that listeners can be trained to attend more to the weak temporal envelope cues to pitch. Some evidence to support this was observed from the results discussed in section 8.1.5 which demonstrated improved discrimination of pitch in complex harmonic stimuli, which included small place variations, after performing
psychophysical rate and modulation rate-pitch discrimination tests. Based on those results and the assumption that specific training with isolated acoustic cues to pitch and spectral timbre might help to reduce the above mentioned problem, a musical-pitch training program was developed.

The training program was designed to primarily teach listeners to attend to F0 information exclusively as a cue to pitch, and secondly to separate that percept from spectral timbre by training on its discrimination using resonant frequency as a cue. It also included subsequent training on discrimination of both pitch and/or spectral timbre using complex sounds in which both F0 and resonant frequency cues varied independently. The program was trialled initially with normal-hearing adults, and later with adult CI recipients using their standard clinical device(s). Outcomes for the NH and CI recipients who used the program demonstrated significant (twofold) benefits to perception of pitch after training as compared to NH and CI control subjects respectively (§7.2.2; §7.3.5.1). Smaller, but significant, benefits to perception of spectral timbre were also observed, but only for the CI recipients (§7.3.5.2). On some of the pitch discrimination tests that were only conducted with the CI users, benefits were observed using stimuli unlike those used during training, and for stimuli which included small variations in resonant frequency. Follow-up tests that were only conducted with the CI recipients showed that pitch discrimination benefits were retained several months after training had ceased. A more detailed discussion of outcomes for the CI recipients follows.

8.2.1 Training with isolated cues

Training with isolated F0 and resonant frequency cues not only improved sensitivity to those cues in stimuli similar to those trained with, but also generalised to improved F0 discrimination in unfamiliar complex stimuli in which both F0 and resonant frequency varied independently. However as variation in the resonant frequencies of those stimuli increased, improvements in F0 discrimination diminished. It would thus seem that whilst adverse effects of place on pitch can be reduced, place variations of the order a few electrode places are too great to be significantly decreased through this form of training. Given that subjects were using their standard clinical CI device(s) during training and testing, which have been shown to have limited capacity to code F0 information, scope for further improvement exists through the additional use of F0
enhancement strategies such as eTone. Further discussion on that topic will follow in section 8.3.

Another outcome worth mentioning is that after training, F0 DLs averaged across subjects and all F0 tests were reduced from approximately 2 to 1 semitone. In practical terms, that absolute level of performance implies that listeners should be able to delineate monophonic tone intervals in music, at least for the range of F0 spanned by those tests.

8.2.2 Subsequent training with multiple cues

No further benefits to F0 or resonant frequency discrimination were observed after additional training with multiple cue variations, compared to results obtained initially through training with isolated cues. In fact, discrimination performance of resonant frequency actually decreased back to a level comparable to that at the onset of the study. The F0 discrimination results on their own suggest either that performance of trainees had saturated during the initial training or that training with multiple cues is ineffective compared to training with isolated cues, at least on those F0 tests. However, the outcomes were somewhat confounded by the fact that subjects trained for a shorter duration of time when training with multiple cues compared to isolated cues. This raises the possibility that further improvements may have resulted had they trained for a comparable period of time. Some evidence to support that argument was presented in section 7.3.8. Further research commencing first with multiple cue training followed by a matched duration of training with isolated cues is required to answer these outstanding issues.

It is interesting to observe that resonant frequency discrimination after additional training with multiple cues variations returned to a performance level that was similar to that obtained prior to any training. That result suggests that the training may have directed subjects’ attention more to F0 cues rather than resonant frequency when listening to musical sounds. That in turn suggests that rather than improve the ability to separate pitch and spectral timbre, the training altered the relative strengths of rate and place cues, and became more biased towards use of rate information.

8.3 Combined F0 enhancement and pitch training

The present research examined two approaches to improve pitch perception in CI recipients, the first involved enhancement of F0 rate cues and the second involved
training to better utilise F0 cues. Both demonstrated benefits and partly alleviated adverse effects of place on pitch (§5.4.3; §7.3.8). While not tested in the present research, it is possible that the combined use of both approaches may provide greater benefit to pitch perception than the use of either approach alone. The study conducted in chapter 5 provided evidence to support that conjecture. It was observed that in addition to the advantage obtained with eTone over ACE, a significant improvement for musically inexperienced subjects was also observed using eTone but not ACE when performance was remeasured after participating in a period of psychophysical rate and modulation rate discrimination tests. Furthermore, given that the psychophysical tests were conducted without feedback, greater improvements may have resulted had directed training been employed. In addition, the two approaches provided different forms of intervention, one enhanced the coding of cues provided at the periphery, and the other presumably operated on processing of cues in higher centres. Assuming that a subject’s performance in discrimination of pitch is not saturated with either approach alone, it seems likely that when combined, their effects may be additive. Unfortunately, it was not possible to test the combined use of eTone and training in the present study because the strategy was not available in a device that could be taken home for an extended period of time.

### 8.4 Conclusions

The experimental rate-pitch coding strategy (eTone), developed to enhance coding of F0 modulation in the stimulus envelope, significantly improved CI recipients’ ranking of monophonic pitch compared to the conventional processing provided by the ACE clinical strategy. The results support the first hypothesis (1a) that more salient and accurate encoding of F0 rate-pitch information, than that provided by existing strategies, can be obtained by employing F0 modulation characterised by a sharp onset and rapid exponential decay, so that each F0 interval is primarily imparted by a single electrical pulse within each channel of stimulation. Improved salience of F0 information was attributed to coding of deep F0 amplitude modulation in phase across channels. While not explicitly measured using the eTone strategy, single electrode psychophysical experiments demonstrated that the pitch-height elicited by the eTone modulation function in a single channel was similar to that produced by an unmodulated pulse train matched in rate to the modulation rate. Given that changes in electrical pulse rate can elicit musical pitch intervals (and probably absolute pitch-heights) that are similar to
Chapter 8: General Discussion and Conclusions

those elicited by F0 variations in acoustic hearing, then the same can be implied for changes in modulation rate of single channel stimuli produced by eTone. Assuming that relationship hold when stimulating many electrodes concurrently, it can be concluded that eTone can convey more accurate musical pitch intervals on the basis of F0 modulation rate than that provided by clinical strategies, at least for low F0s up to approximately 300 Hz.

The improved pitch ranking results using eTone were greatest for listeners with no formal music experience and after gaining experience attending to rate information as a cue to pitch, suggesting that rate-pitch training in a psychophysical context can improve F0 discrimination in more complex sounds. While performance of CI users improved using eTone compared to ACE, it remained poorer than that of normal-hearing listeners with comparable musical experience. In addition, adverse effects of small variations in formant frequencies amongst the stimuli on ranking of pitch were observed, mainly when using ACE by all listeners, and only by musically naïve listeners using eTone. These results, together with those for similar effects of place of stimulation on rate-pitch observed for poorer-performing subjects in the single electrode psychophysical tests, highlight the problem that musically inexperienced CI listeners face when judging the pitch of complex sounds.

No differences in speech recognition either in quiet or in noise were observed between eTone and ACE. Although the deep modulation applied by eTone for harmonic sounds can result in reduced loudness compared to ACE, it was compensated for by increasing gain in F0 modulated channels. These increases in gain expressed in terms of changes to stimulation levels were consistent with those predicted by psychophysics tests comparing the loudness of unmodulated and modulated pulse trains presented on a single electrode. Given that distortion of channel loudness can affect coding of spectral and low-frequency temporal envelope information, and hence speech recognition, the appropriateness of the loudness compensation in eTone may be reflected in the comparable speech recognition between strategies. These results support hypothesis (1b) that compensation for effects on loudness produced by application of deep F0 modulation may be needed so as to produce similar loudness to that produced by existing strategies and thereby maintain existing levels of speech perception. The comparable speech perception results between strategies coupled with the improved coding of F0 pitch suggest that recognition of lexical tone in tonal languages can also be improved using eTone.
Significant improvements in discrimination of F0, and to a lesser extent resonant frequency were observed for CI recipients after training with a musical pitch training program. In contrast, no improvement in those discrimination measures was seen for control CI recipients that did not partake in the training. The results support the hypothesis (2a) that training involving discrimination of single cues to attributes of musical sounds (i.e., F0 and resonant frequency) in the absence of other cue variations will improve listeners’ sensitivity to those cues. Some of the results lend support to the second part of the hypothesis (2b) that training with single isolated cues can generalise to improved F0 discrimination in more complex real-world sounds in which multiple cues vary. However, little improvement was seen when large variations in place cues were present. Those results suggest that adverse effects of place of stimulation on subjects’ judgement of pitch can be reduced by the training for small but not large variations in place. Assessment of the third part of the hypothesis (2c), that subsequent training with complex sounds in which multiple cues vary may further improve listeners’ ability to discriminate the pitch of complex sounds, was inconclusive, due mainly to issues related to saturation of performance during the initial training phase and differences in training duration across training phases. Follow-up evaluations with trainees conducted well after training had ceased verified the robustness of the training effect on F0 discrimination but not resonant frequency discrimination.

The overall results of the training study were very encouraging as demonstrated by the twofold improvement in pitch discrimination of complex harmonic tones unlike those trained with, and smaller but significant improvements for unfamiliar harmonic tones that included small variations in spectral timbre. Absolute levels of pitch perception by the CI recipients improved to the point in which average discrimination thresholds were around 1 semitone. In practical terms those improvements suggest that such training techniques, particularly the training with isolated single cues, can help CI users to better delineate pitch intervals in music and thereby improve their ability to follow melody. Furthermore, while not investigated in the present research, results did suggest that greater improvements to pitch discrimination in complex sounds compared to those observed in the present training study may be possible if the training is coupled with use of an enhanced rate-pitch coding strategy such as eTone.
Bibliography

Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


248

Appendices

A. aTune rewards

- **Stop**: stops the movement of all Tokens on the screen for a short period of time (typically 10 secs).

- **Reverse**: reverses the movement of all Tokens on the screen.

- **Blast**: blows-up the leading most section of Tokens on the screen. Typically up to 8 Token Patterns are eliminated.

- **Zap**: when used, the cursor will change to a lighting bolt and if the next response is correct, then up to 8 identical Patterns on the screen will also be eliminated.

- **Hint**: the correct Pattern of Tokens will be highlighted in aqua-blue colour for a short period of time (typically 10 secs).

- **Slow**: slows the progression of all Tokens for a short period of time (typically 20 secs).

- **Hammer**: when used, the cursor will change to a hammer for a short period of time (typically 4 secs) allowing the trainee to bash any Tokens on the screen to eliminate them.

- **Cycle Target**: when used, a 60 second period will ensure in which the audible/acoustic tone Pattern can be cycled to another random Pattern by pressing the space bar or by pressing the left-mouse button when the cursor is positioned on an empty part of the screen where there are no visual Tokens.
### B. aTune Level and Stage details

#### Stage 1a – Pitch only

<table>
<thead>
<tr>
<th>Level</th>
<th>Pattern</th>
<th>Num Timbres</th>
<th>Num Sequential Tones</th>
<th>Num Combs</th>
<th>Expected Response Time (secs)</th>
<th>Response Time + Presentation Time (secs)</th>
<th>Patterns Per Block</th>
<th>New Block Delay (secs)</th>
<th>Token Velocity Scalar</th>
<th>Fast Response Time (secs)</th>
<th>Resonant Freq Spectral Envelope Bandwidth</th>
<th>Temporal Envelope Shape</th>
<th>Level Jitter (dB)</th>
<th>Correct Pattern Insertion Feedback</th>
<th>Feedback Path</th>
<th>Rewards</th>
<th>Num Jewels</th>
<th>Num Obstacles</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0.8</td>
<td>1.4</td>
<td>8</td>
<td>11</td>
<td>1.2</td>
<td>2</td>
<td>Low / Narrow</td>
<td>Uniform</td>
<td>0</td>
<td>Yes</td>
<td>Pipes</td>
<td>Auto Reverse</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>1.1</td>
<td>2.2</td>
<td>8</td>
<td>18</td>
<td>1.2</td>
<td>3</td>
<td>Mid / Narrow</td>
<td>Uniform</td>
<td>0</td>
<td>Yes</td>
<td>Pipes</td>
<td>Stop</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>1.1</td>
<td>2.2</td>
<td>8</td>
<td>18</td>
<td>1.4</td>
<td>3</td>
<td>Mid / Wide</td>
<td>Uniform</td>
<td>3</td>
<td>Yes</td>
<td>Void Borders</td>
<td>Stop</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>1.1</td>
<td>2.2</td>
<td>8</td>
<td>18</td>
<td>1.6</td>
<td>3</td>
<td>Low / Narrow</td>
<td>Uniform</td>
<td>3</td>
<td>Yes</td>
<td>Back and Forth</td>
<td>Reverse</td>
<td>5</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>1.1</td>
<td>2.2</td>
<td>8</td>
<td>18</td>
<td>1.8</td>
<td>3</td>
<td>Mid / Narrow</td>
<td>Uniform</td>
<td>6</td>
<td>Yes</td>
<td>Ramps</td>
<td>Blast</td>
<td>5</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>1.1</td>
<td>2.2</td>
<td>8</td>
<td>18</td>
<td>2</td>
<td>2</td>
<td>High / Narrow</td>
<td>Swell-up</td>
<td>6</td>
<td>Yes</td>
<td>One Bounce</td>
<td>Zap</td>
<td>5</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>1.0</td>
<td>1.6</td>
<td>10</td>
<td>16</td>
<td>1.6</td>
<td>2</td>
<td>Mid / Wide</td>
<td>Swell-up</td>
<td>6</td>
<td>No</td>
<td>Random Mix</td>
<td>Hammer</td>
<td>10</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>1.0</td>
<td>1.6</td>
<td>10</td>
<td>16</td>
<td>2</td>
<td>2</td>
<td>Mid / Narrow</td>
<td>Sharp-onset</td>
<td>6</td>
<td>Yes</td>
<td>Ramps</td>
<td>Cycle Target</td>
<td>15</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>2</td>
<td>9</td>
<td>1.9</td>
<td>3.0</td>
<td>6</td>
<td>18</td>
<td>1.6</td>
<td>3</td>
<td>High / Wide</td>
<td>Fade-out</td>
<td>6</td>
<td>No</td>
<td>Random Mix</td>
<td>Blast, Zap, Slow</td>
<td>15</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1.0</td>
<td>1.6</td>
<td>10</td>
<td>16</td>
<td>2</td>
<td>2</td>
<td>Mid / Narrow</td>
<td>Sharp-onset</td>
<td>6</td>
<td>Yes</td>
<td>One Bounce</td>
<td>All</td>
<td>10</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>1.9</td>
<td>3.0</td>
<td>6</td>
<td>18</td>
<td>1.5</td>
<td>3</td>
<td>Mid / Wide</td>
<td>Fade-out</td>
<td>6</td>
<td>No</td>
<td>Random Voids</td>
<td>All</td>
<td>15</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>1.9</td>
<td>3.0</td>
<td>6</td>
<td>18</td>
<td>1.5</td>
<td>3</td>
<td>High / Wide</td>
<td>Swell-up</td>
<td>6</td>
<td>No</td>
<td>Round and Round</td>
<td>All</td>
<td>20</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

#### Stage 1b – Spectral Timbre only

<table>
<thead>
<tr>
<th>Level</th>
<th>Pattern</th>
<th>Num Timbres</th>
<th>Num Sequential Tones</th>
<th>Num Combs</th>
<th>Expected Response Time (secs)</th>
<th>Response Time + Presentation Time (secs)</th>
<th>Patterns Per Block</th>
<th>New Block Delay (secs)</th>
<th>Token Velocity Scalar</th>
<th>Fast Response Time (secs)</th>
<th>Resonant Freq Spectral Envelope Bandwidth</th>
<th>Temporal Envelope Shape</th>
<th>Level Jitter (dB)</th>
<th>Correct Pattern Insertion Feedback</th>
<th>Feedback Path</th>
<th>Rewards</th>
<th>Num Jewels</th>
<th>Num Obstacles</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0.8</td>
<td>1.4</td>
<td>8</td>
<td>11</td>
<td>1.4</td>
<td>2</td>
<td>Low / Narrow</td>
<td>Uniform</td>
<td>0</td>
<td>Yes</td>
<td>Pipes</td>
<td>Auto Reverse</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>1.1</td>
<td>2.2</td>
<td>8</td>
<td>18</td>
<td>1.4</td>
<td>3</td>
<td>Mid / Narrow</td>
<td>Uniform</td>
<td>0</td>
<td>Yes</td>
<td>Pipes</td>
<td>Stop, Zap, Hint</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>1.1</td>
<td>2.2</td>
<td>8</td>
<td>18</td>
<td>1.5</td>
<td>3</td>
<td>Mid / Wide</td>
<td>Uniform</td>
<td>3</td>
<td>Yes</td>
<td>Void Borders</td>
<td>Rev, Blast, Zap</td>
<td>5</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>1.1</td>
<td>2.2</td>
<td>8</td>
<td>18</td>
<td>1.6</td>
<td>3</td>
<td>Low / Wide</td>
<td>Sharp-onset</td>
<td>6</td>
<td>Yes</td>
<td>Ramps</td>
<td>All</td>
<td>5</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>1.1</td>
<td>2.2</td>
<td>8</td>
<td>18</td>
<td>1.8</td>
<td>2</td>
<td>High / Narrow</td>
<td>Swell-up</td>
<td>6</td>
<td>Yes</td>
<td>Holes</td>
<td>All</td>
<td>5</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>1.1</td>
<td>2.2</td>
<td>8</td>
<td>18</td>
<td>2</td>
<td>2</td>
<td>Mid / Wide</td>
<td>Fade-out</td>
<td>6</td>
<td>Yes</td>
<td>One Bounce</td>
<td>All</td>
<td>10</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>3</td>
<td>8</td>
<td>1.7</td>
<td>3.3</td>
<td>6</td>
<td>20</td>
<td>1.6</td>
<td>4</td>
<td>High / Wide</td>
<td>Fade-out</td>
<td>6</td>
<td>No</td>
<td>Random Mix</td>
<td>All</td>
<td>10</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>3</td>
<td>8</td>
<td>1.7</td>
<td>3.3</td>
<td>6</td>
<td>20</td>
<td>1.6</td>
<td>4</td>
<td>High / Wide</td>
<td>Fade-out</td>
<td>6</td>
<td>No</td>
<td>Random Mix</td>
<td>All</td>
<td>10</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1.0</td>
<td>1.6</td>
<td>10</td>
<td>16</td>
<td>1.6</td>
<td>2</td>
<td>Mid / Wide</td>
<td>Swell-up</td>
<td>6</td>
<td>No</td>
<td>Two Bounce</td>
<td>All</td>
<td>10</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1.0</td>
<td>1.6</td>
<td>10</td>
<td>16</td>
<td>1.5</td>
<td>2</td>
<td>Mid / Narrow</td>
<td>Uniform</td>
<td>6</td>
<td>No</td>
<td>Round and Round</td>
<td>All</td>
<td>15</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>1.9</td>
<td>3.0</td>
<td>6</td>
<td>18</td>
<td>1.5</td>
<td>3</td>
<td>Mid / Wide</td>
<td>Fade-out</td>
<td>6</td>
<td>No</td>
<td>Round and Round</td>
<td>All</td>
<td>15</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>1.9</td>
<td>3.0</td>
<td>6</td>
<td>18</td>
<td>2</td>
<td>3</td>
<td>High / Wide</td>
<td>Swell-up</td>
<td>6</td>
<td>No</td>
<td>Round and Bounce</td>
<td>All</td>
<td>20</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>
### Stage 2a – Combined Pitch and Spectral Timbre

<table>
<thead>
<tr>
<th>Level</th>
<th>Num F0s</th>
<th>Num Timbres</th>
<th>Num Sequential Tones</th>
<th>Num Combs</th>
<th>Expected Response Time (secs)</th>
<th>Response Time + Presentation Time (secs)</th>
<th>Patterns Per Block</th>
<th>New Block Delay (secs)</th>
<th>Token Velocity Scalar</th>
<th>Fast Response Time (secs)</th>
<th>Fat Response Time (secs)</th>
<th>Resonant Freq / Spectral Envelope Bandwidth</th>
<th>Temporal Envelope Shape</th>
<th>Level Filter (dB)</th>
<th>Correct Pattern Insertion Feedback</th>
<th>Path</th>
<th>Rewards</th>
<th>Num Jewels</th>
<th>Num Obstacles</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2*</td>
<td>0.8</td>
<td>1.4</td>
<td>10</td>
<td>14</td>
<td>1.3</td>
<td>2</td>
<td>Mid / Narrow</td>
<td>Uniform</td>
<td>Yes</td>
<td>Pipes</td>
<td>Auto Reverse</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>4*</td>
<td>1.1</td>
<td>2.2</td>
<td>8</td>
<td>18</td>
<td>1.5</td>
<td>3</td>
<td>Mid / Wide</td>
<td>Sharp Onset</td>
<td>No</td>
<td>Void Borders</td>
<td>Auto Reverse</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>1.1</td>
<td>2.2</td>
<td>8</td>
<td>18</td>
<td>1.6</td>
<td>3</td>
<td>Low / Narrow</td>
<td>Swell-up</td>
<td>No</td>
<td>Back and Forth</td>
<td>Auto Reverse</td>
<td>5</td>
<td>0</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>1.1</td>
<td>2.2</td>
<td>8</td>
<td>14</td>
<td>1.2</td>
<td>3</td>
<td>Mid / Wide</td>
<td>Uniform</td>
<td>Yes</td>
<td>Ramps</td>
<td>All</td>
<td>5</td>
<td>0</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>16</td>
<td>2.9</td>
<td>4.0</td>
<td>5</td>
<td>20</td>
<td>1.2</td>
<td>4</td>
<td>Mid / Narrow</td>
<td>Swell-up</td>
<td>Yes</td>
<td>Holes</td>
<td>All</td>
<td>5</td>
<td>0</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>16</td>
<td>2.9</td>
<td>4.0</td>
<td>5</td>
<td>20</td>
<td>1.6</td>
<td>4</td>
<td>Low / Narrow</td>
<td>Fade-out</td>
<td>No</td>
<td>One Bounce</td>
<td>All</td>
<td>10</td>
<td>1</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>6</td>
<td>1.4</td>
<td>2.0</td>
<td>8</td>
<td>16</td>
<td>1.6</td>
<td>3</td>
<td>High / Narrow</td>
<td>Swell-up</td>
<td>Yes</td>
<td>Random Mix</td>
<td>All</td>
<td>10</td>
<td>1</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>36</td>
<td>5.9</td>
<td>7.0</td>
<td>4</td>
<td>28</td>
<td>1.2</td>
<td>6</td>
<td>Mid / Wide</td>
<td>Sharp Onset</td>
<td>Yes</td>
<td>Two Bounce</td>
<td>All</td>
<td>10</td>
<td>1</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>36</td>
<td>5.9</td>
<td>7.0</td>
<td>4</td>
<td>28</td>
<td>1.6</td>
<td>6</td>
<td>High / Wide</td>
<td>Uniform</td>
<td>No</td>
<td>Round and Round</td>
<td>All</td>
<td>15</td>
<td>1</td>
<td>15</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>36</td>
<td>5.9</td>
<td>7.0</td>
<td>4</td>
<td>28</td>
<td>1.6</td>
<td>6</td>
<td>Low / Narrow</td>
<td>Fade-out</td>
<td>No</td>
<td>Random Voids</td>
<td>All</td>
<td>15</td>
<td>2</td>
<td>15</td>
<td>2</td>
</tr>
<tr>
<td>11</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>9</td>
<td>1.9</td>
<td>2.5</td>
<td>6</td>
<td>15</td>
<td>1.2</td>
<td>3</td>
<td>Mid / Wide</td>
<td>Fade-out</td>
<td>Yes</td>
<td>Up and Down</td>
<td>All</td>
<td>15</td>
<td>2</td>
<td>15</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>9</td>
<td>1.9</td>
<td>2.5</td>
<td>6</td>
<td>15</td>
<td>1.6</td>
<td>3</td>
<td>Mid / Narrow</td>
<td>Swell-up</td>
<td>No</td>
<td>Round and Bounce</td>
<td>All</td>
<td>20</td>
<td>2</td>
<td>20</td>
<td>2</td>
</tr>
</tbody>
</table>

### Stage 2b – Instruments: Pitch and Timbre

<table>
<thead>
<tr>
<th>Level</th>
<th>Num F0s</th>
<th>Num Timbres</th>
<th>Num Sequential Tones</th>
<th>Num Combs</th>
<th>Expected Response Time (secs)</th>
<th>Response Time + Presentation Time (secs)</th>
<th>Patterns Per Block</th>
<th>New Block Delay (secs)</th>
<th>Token Velocity Scalar</th>
<th>Fast Response Time (secs)</th>
<th>Fat Response Time (secs)</th>
<th>Instrument or Vowel Level Jitter (dB)</th>
<th>Correct Pattern Insertion Feedback</th>
<th>Path</th>
<th>Rewards</th>
<th>Num Jewels</th>
<th>Num Obstacles</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>1.1</td>
<td>1.7</td>
<td>10</td>
<td>17</td>
<td>1.6</td>
<td>2</td>
<td>Tromb, Piano</td>
<td>6</td>
<td>Yes</td>
<td>Pipes</td>
<td>All</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>16</td>
<td>2.9</td>
<td>4.0</td>
<td>6</td>
<td>24</td>
<td>1.6</td>
<td>4</td>
<td>Tromb, Piano</td>
<td>6</td>
<td>Yes</td>
<td>Void Borders</td>
<td>All</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>16</td>
<td>2.9</td>
<td>4.0</td>
<td>6</td>
<td>24</td>
<td>1.8</td>
<td>3</td>
<td>Tromb, Piano</td>
<td>6</td>
<td>Yes</td>
<td>Back and Forth</td>
<td>All</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>16</td>
<td>2.9</td>
<td>4.0</td>
<td>5</td>
<td>20</td>
<td>1.2</td>
<td>3</td>
<td>/a/, /i/</td>
<td>6</td>
<td>No</td>
<td>Ramps</td>
<td>All</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>16</td>
<td>2.9</td>
<td>4.0</td>
<td>5</td>
<td>20</td>
<td>1.2</td>
<td>3</td>
<td>/a/, /i/</td>
<td>6</td>
<td>No</td>
<td>Holes</td>
<td>All</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>6</td>
<td>1.4</td>
<td>2.0</td>
<td>10</td>
<td>20</td>
<td>1.6</td>
<td>2</td>
<td>/a/, /u/, /i/</td>
<td>6</td>
<td>No</td>
<td>Once Bounce</td>
<td>All</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>36</td>
<td>5.9</td>
<td>7.0</td>
<td>5</td>
<td>35</td>
<td>1.6</td>
<td>6</td>
<td>Tromb, Sax, Piano</td>
<td>6</td>
<td>No</td>
<td>Random Mix</td>
<td>All</td>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>6</td>
<td>1.4</td>
<td>2.0</td>
<td>10</td>
<td>20</td>
<td>2</td>
<td>2</td>
<td>Tromb, Sax, Piano</td>
<td>6</td>
<td>No</td>
<td>Two Bounce</td>
<td>All</td>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>36</td>
<td>5.9</td>
<td>7.0</td>
<td>5</td>
<td>35</td>
<td>1.6</td>
<td>6</td>
<td>Tromb, Sax, Piano</td>
<td>6</td>
<td>No</td>
<td>Round and Round</td>
<td>All</td>
<td>15</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>36</td>
<td>5.9</td>
<td>7.0</td>
<td>5</td>
<td>35</td>
<td>1.6</td>
<td>6</td>
<td>Tromb, Sax, Piano</td>
<td>6</td>
<td>No</td>
<td>Random Voids</td>
<td>All</td>
<td>15</td>
<td>2</td>
</tr>
<tr>
<td>11</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>9</td>
<td>1.9</td>
<td>2.5</td>
<td>8</td>
<td>20</td>
<td>1.8</td>
<td>3</td>
<td>Tromb, Sax, Piano</td>
<td>6</td>
<td>Yes</td>
<td>Round and Bounce</td>
<td>All</td>
<td>20</td>
<td>3</td>
</tr>
<tr>
<td>12</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>81</td>
<td>12.7</td>
<td>13.8</td>
<td>4</td>
<td>55</td>
<td>1.4</td>
<td>12</td>
<td>Tromb, Sax, Piano</td>
<td>6</td>
<td>Yes</td>
<td>Round and Bounce</td>
<td>All</td>
<td>20</td>
<td>3</td>
</tr>
</tbody>
</table>
### Stage 2c – Instruments: Pitch only

<table>
<thead>
<tr>
<th>Level</th>
<th>NumF0s</th>
<th>Num Timbres</th>
<th>Num Sequential Tones</th>
<th>Num Combs</th>
<th>Expected Response Time (secs)</th>
<th>Patterns Per Block</th>
<th>New Block Delay (secs)</th>
<th>Token Velocity</th>
<th>Fast Response Time (secs)</th>
<th>Instrument or Vowel Level</th>
<th>Jitter (dB)</th>
<th>Correct Pattern Insertion Feedback</th>
<th>Path</th>
<th>Rewards</th>
<th>Num Jewels</th>
<th>Num Obstacles</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>1.1</td>
<td>2.2</td>
<td>8</td>
<td>18</td>
<td>1.6</td>
<td>Tromb, Piano</td>
<td>6</td>
<td>Yes</td>
<td>Pipes</td>
<td>All</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>1.1</td>
<td>2.2</td>
<td>8</td>
<td>18</td>
<td>1.6</td>
<td>Tromb, Piano</td>
<td>6</td>
<td>No</td>
<td>Void Borders</td>
<td>All</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>1.1</td>
<td>2.2</td>
<td>8</td>
<td>18</td>
<td>1.8</td>
<td>Tromb, Piano</td>
<td>6</td>
<td>Yes</td>
<td>Back and Forth</td>
<td>All</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>1.1</td>
<td>2.2</td>
<td>8</td>
<td>18</td>
<td>2</td>
<td>/i/, /b/</td>
<td>6</td>
<td>No</td>
<td>Ramps</td>
<td>All</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>8</td>
<td>1.7</td>
<td>3.3</td>
<td>6</td>
<td>20</td>
<td>1.6</td>
<td>/i/, /b/</td>
<td>6</td>
<td>No</td>
<td>Holes</td>
<td>All</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>1.1</td>
<td>2.2</td>
<td>8</td>
<td>18</td>
<td>1.8</td>
<td>/i/, /u/, /i/</td>
<td>6</td>
<td>No</td>
<td>Once Bounce</td>
<td>All</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>1.1</td>
<td>2.2</td>
<td>8</td>
<td>18</td>
<td>2</td>
<td>Tromb, Sax, Piano</td>
<td>6</td>
<td>No</td>
<td>Random Mix</td>
<td>All</td>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>8</td>
<td>1.7</td>
<td>3.3</td>
<td>6</td>
<td>20</td>
<td>1.6</td>
<td>Tromb, Sax, Piano</td>
<td>6</td>
<td>No</td>
<td>Two Bounce</td>
<td>All</td>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>8</td>
<td>1.7</td>
<td>3.3</td>
<td>6</td>
<td>20</td>
<td>1.8</td>
<td>/i/, /u/, /i/</td>
<td>6</td>
<td>No</td>
<td>Round and Round</td>
<td>All</td>
<td>15</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>9</td>
<td>1.9</td>
<td>3.0</td>
<td>6</td>
<td>18</td>
<td>1.8</td>
<td>Tromb, Sax, Piano</td>
<td>6</td>
<td>No</td>
<td>Random Voids</td>
<td>All</td>
<td>15</td>
<td>2</td>
</tr>
<tr>
<td>11</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>9</td>
<td>1.9</td>
<td>3.0</td>
<td>7</td>
<td>21</td>
<td>1.8</td>
<td>/i/, /u/, /i/</td>
<td>6</td>
<td>No</td>
<td>Up and Down</td>
<td>All</td>
<td>15</td>
<td>3</td>
</tr>
<tr>
<td>12</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>27</td>
<td>4.6</td>
<td>6.2</td>
<td>4</td>
<td>25</td>
<td>1.6</td>
<td>Tromb, Sax, Piano</td>
<td>6</td>
<td>No</td>
<td>Round and Bounce</td>
<td>All</td>
<td>20</td>
<td>3</td>
</tr>
</tbody>
</table>

Table B.1. Specific details for training/program parameters used in each Level of each Stage of the training game.

* the number of combinations for those denoted Levels was determined from tones in which only consistent cues to pitch were provided by variations in F0 and resonant frequencies, thus halving the total number of combinations.
C. aTune Paths (screen configurations)

1. Pipes

2. Back and Forth

3. Ramps

4. Holes

5. Void Borders

6. Random Voids
Appendices

7. Random Mix

8. Up and Down

9. One Bounce

10. Two Bounce

11. Round and Round

12. Round and Bounce

Figure C.1. Screen configurations for each path type. The entry point (start) of the path is depicted by the green arrow and the exit (end) by the red arrow. Blue arrows highlight the path that visual Tokens take when travelling in their forward direction. Purple dashed arrows depict an invisible path between two points that the visual Tokens are transported between. For the Ramps and Holes paths, the locations of the ramps and holes respectively vary from one Run to the next. Similarly, for the Void Borders and Random Voids paths, the paths and location of voids vary between Runs. For the Random Mix path, ramps, holes, or voids which vary in location are randomly allocated between Runs.
Author/s: 
Vandali, Andrew E.

Title: 
Optimisation of rate-pitch perception in cochlear implant hearing

Date: 
2013

Citation: 

Persistent Link: 
http://hdl.handle.net/11343/39956

File Description: 
Optimisation of rate-pitch perception in cochlear implant hearing

Terms and Conditions: 
Terms and Conditions: Copyright in works deposited in Minerva Access is retained by the copyright owner. The work may not be altered without permission from the copyright owner. Readers may only download, print and save electronic copies of whole works for their own personal non-commercial use. Any use that exceeds these limits requires permission from the copyright owner. Attribution is essential when quoting or paraphrasing from these works.