Noise reduction for cochlear implants

Adam Anthony HERSBACH
Submitted in partial fulfilment of the requirements of the degree of Doctor of Philosophy (with coursework component)

April 2015
Department of Electronic and Electrical Engineering
University of Melbourne
Abstract

Cochlear implant (CI) users generally achieve acceptable speech understanding in quiet conditions, but have difficulty understanding speech in the presence of background noise. In this case, noise reduction processing can be utilised to help improve the situation, and solutions can be distinguished based on the number of microphones used to sample the acoustic environment. Single microphone solutions rely on the statistical properties of speech and noise while multi-microphone solutions can use the spatial characteristics of impinging sound to further separate speech from noise. It is the latter that forms the focus of the current research.

A multi-microphone noise reduction algorithm was developed for a CI sound processor that attenuated sound from the rear while passing sound from in front of the listener. The algorithm used two microphones with small physical separation to generate two fixed directional patterns; one facing forward, the other towards the rear. By examining the front-to-back energy ratio, a signal-to-noise ratio (SNR) estimate was obtained, which was used to attenuate noise dominated frequency channels.

The algorithm was evaluated acutely with CI listeners, primarily using an adaptive speech reception threshold (SRT) task, although sound quality and acceptable noise level were also studied. The acoustic environment used for evaluation in the laboratory included complex situations. These situations used various numbers of competing talkers or interfering speech weighted noise sources that changed spatial locations during the test. Reverberation was introduced and the algorithm was evaluated in a range of reverberant environments. Microphone sensitivity matching was investigated by introducing controlled levels of mismatch and measuring speech intelligibility performance.

The evaluation revealed the algorithm was highly beneficial across a wide range of acoustic situations, outperforming a conventional generalised side-lobe canceller algorithm called Beam. The benefit varied with the spatial configuration of the competing noise and was greatest when the noise was located to the sides and rear of the listener. The benefit in reverberant conditions was maintained. Counter-intuitively, the benefits actually increased in the highest level of reverberation that was evaluated. Microphone mismatch had a detrimental effect on all multi-microphone algorithms that were evaluated, completely negating any multi-microphone benefit when the mismatch was 4 dB or greater.
Finally, the algorithm was implemented in a wearable sound processor and CI users evaluated the algorithm outside the laboratory during their normal use of the device. Users were able to vote for their preferred listening program using their processor’s remote control device. The take-home evaluation consolidated the benefits measured acutely in the laboratory and provides critical guidance as to how the algorithm could be integrated into a commercial device.
Declaration

This is to certify that

- the thesis comprises only my original work towards the PhD except where indicated in the Preface,
- due acknowledgement has been made in the text to all other material used,
- the thesis is fewer than 100 000 words in length, exclusive of tables, maps, bibliographies and appendices.

____________________________________________

Adam Hersbach
Preface

This work was financially supported by Cochlear Ltd where Adam was employed full time while studying his PhD. He undertook all work whilst located at the Cochlear laboratory in East Melbourne. There were many people from the lab as well as lab-resources involved in the work undertaken as part of this PhD. In particular, there was essential support for clinical evaluation of cochlear implant (CI) users in the sound booth. This included co-operation of both the audiological staff as well as the use of sound booth facilities. These were required to undertake the extensive clinical evaluation and speech testing with CI users.

The experiments on high reverberation (Chapter 6) and microphone mismatch (Chapter 7) were undertaken by Michael Segal and Nadirah Mannan, respectively, during research projects conducted at Cochlear Ltd as University of Melbourne Master of Audiology students. Lisa Dang, also a Master of Audiology student, studied normal hearing performance that contributed to the data for normal hearers provided in Chapter 8. The students were given the experimental objectives and clinical protocols at the commencement of the program, so were not involved in the design of the experiments. Their involvement was primarily with collection of data during speech testing in the sound booth which the students also used for write-up and analysis in their Masters projects. Michael Segal made additional contributions by configuring and using the room simulator software to generate reverberant impulse responses used in Chapter 6. All data analysis and discussion provided in this thesis are Adam’s own work.

Portions of this work were published during Adam’s PhD candidature; in particular, data from experiments in Chapters 4 and 5 have been published in Hersbach et al. (2013a) and Hersbach et al. (2013b) and are available as appendices. Data from Chapter 6 was accepted for publication at the International Conference on Acoustics, Speech and Signal Processing (ICASSP) April 2015, and is also available in full in the appendix.
Acknowledgments

Cochlear Ltd provided the financial support for this project. I was employed full-time throughout the duration of my PhD candidature and I thank Jim Patrick of Cochlear Ltd for supporting this opportunity. Throughout the project, I was located at the Cochlear laboratory in East Melbourne where the experiments were primarily conducted. As such, many colleagues from Cochlear were involved in the research project, and I am indebted to them and am thankful for their support in completing this thesis.

I would like to thank Stefan Mauger and Pam Dawson of Cochlear Ltd for discussions that provided many insightful and interesting insights on predictive metrics associated with error rates and noise reduction algorithms. I would also like to thank Pam Dawson for the time she spent providing statistical advice, particularly in regard to teaching the art rather than simply using the tools. I would like to thank Jennie Gorrie, Komal Arora, Michelle Knight, Ruth English and Chris Warren of Cochlear Ltd for their assistance and advice on clinical protocol design and for the many hours they spent in the sound booth collecting data.

Three Master of Audiology students from University of Melbourne were involved with parts of the project under supervision. Michael Segal utilised the MCRoomSim toolbox to produce the input stimuli for the experiment in Chapter 6 and also undertook data collection. Nadirah Mannan undertook data collection in Chapter 7 and Lisa Dang collected data for the normal hearing benchmark group in Chapter 8.

From National Acoustics Laboratory (NAL), Jorge Mejia and Chris Orinos made the impulse response recordings in the reverberant training room at NAL that were used in Chapters 5 and 6. Richard van Hoesel of CRC Hear advised on the choice of noise spatial configurations. I would like to acknowledge the Cochlear Ltd DSP team in Sydney for the firmware implementation of the noise reduction algorithm on the CP900 behind-the-ear (BTE) sound processor used in Chapter 8.

Bas van Dijk and Michael Goorevich of Cochlear Ltd, through management of the Cochlear Sound Coding Research network, have been intimately involved in the evolution of the noise reduction algorithm, and have enabled resources within Cochlear required to complete this work.

Kerrie Plant and John Heasman of Cochlear Ltd held management roles within the Cochlear Melbourne office and supported this work extensively. I thank Kerrie for advice on clinical protocol design and for devoting scarce audiological resources to the project, and to John for
Acknowledgments

providing the perfect environment to undertake this work. I also thank John for providing the means to establish the technical tools including xPC real time hardware, sound booth, loudspeakers and supporting acoustic equipment required to carry out the experiments efficiently and reliably.

My PhD supervisors Hugh McDermott, David Grayden and James Fallon always pushed for the best outcomes and for this I thank them. Their review during the manuscript drafting process was immensely beneficial, and I thank them for the time and attention they paid to various draft documents - the completed work was greatly enhanced through their advice. My advisory committee members, Pam Dawson and Richard van Hoesel, also steered the project in the right direction, and encouraged critical thinking across different aspects of the project.

I would also like to sincerely thank my father, Tony Hersbach, who provided valuable editorial corrections for the final draft. Thanks for reading it cover to cover!

Not least, I would thank the 33 different research volunteers who graciously donated their time and effort to endure hours upon hours of speech testing in the sound booth. Without them, this work would not have been possible.

Finally, I dedicate this work to my beautiful wife Megan and our three wonderful children, Darcy, Quinn and Willow, who asked: ‘Is that all your experiments, Dad?’ ‘Is that what you typed on the computer, Dad?’ and ‘Where’s the pictures, Dad?’
### Glossary

#### Acronyms and abbreviations

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>20-TB-full</td>
<td>Twenty talker babble noise, full circle presentation</td>
</tr>
<tr>
<td>2-SWN-rear</td>
<td>Two speech weighted noise, rear hemi-field presentation</td>
</tr>
<tr>
<td>4-TB-full</td>
<td>Four talker babble noise, full circle presentation</td>
</tr>
<tr>
<td>4-TB-rear</td>
<td>Four talker babble noise, rear hemi-field presentation</td>
</tr>
<tr>
<td>ACE</td>
<td>Advanced combination encoder</td>
</tr>
<tr>
<td>ADM</td>
<td>Adaptive directional microphone</td>
</tr>
<tr>
<td>ADRO</td>
<td>Adaptive dynamic range optimisation</td>
</tr>
<tr>
<td>AI</td>
<td>Articulation index</td>
</tr>
<tr>
<td>ANL</td>
<td>Acceptable noise level</td>
</tr>
<tr>
<td>ANOVA</td>
<td>Analysis of variance</td>
</tr>
<tr>
<td>APHAB</td>
<td>Abbreviated profile of hearing aid benefit</td>
</tr>
<tr>
<td>AuSTIN</td>
<td>Australian sentence test in noise</td>
</tr>
<tr>
<td>BKB</td>
<td>Bamford-Knowles-Bench</td>
</tr>
<tr>
<td>BTE</td>
<td>Behind-the-ear</td>
</tr>
<tr>
<td>C.I.</td>
<td>95% confidence interval</td>
</tr>
<tr>
<td>C50</td>
<td>Clarity (50 ms)</td>
</tr>
<tr>
<td>C7</td>
<td>Clarity (7 ms)</td>
</tr>
<tr>
<td>CI</td>
<td>Cochlear implant</td>
</tr>
<tr>
<td>CIS</td>
<td>Continuous interleaved sampling</td>
</tr>
<tr>
<td>CP900</td>
<td>Cochlear Nucleus 6 sound processor</td>
</tr>
<tr>
<td>CR230</td>
<td>Cochlear Nucleus 6 remote control</td>
</tr>
<tr>
<td>C-SPL</td>
<td>Comfort-level sound pressure level</td>
</tr>
<tr>
<td>dB</td>
<td>Decibel</td>
</tr>
<tr>
<td>dB SPL</td>
<td>Decibel sound pressure level</td>
</tr>
<tr>
<td>DNN</td>
<td>Deep neural network</td>
</tr>
<tr>
<td>DOA</td>
<td>Direction of arrival</td>
</tr>
<tr>
<td>DRR</td>
<td>Direct-to-reverberant ratio</td>
</tr>
<tr>
<td>EDT</td>
<td>Early decay time</td>
</tr>
<tr>
<td>FA</td>
<td>False alarm rate</td>
</tr>
<tr>
<td>FFT</td>
<td>Fast Fourier transform</td>
</tr>
<tr>
<td>GMM</td>
<td>Gaussian mixture model</td>
</tr>
<tr>
<td>HINT</td>
<td>Hearing in noise test</td>
</tr>
<tr>
<td>HIT</td>
<td>Hit rate</td>
</tr>
<tr>
<td>Hz</td>
<td>Hertz</td>
</tr>
<tr>
<td>IIR</td>
<td>Infinite impulse response</td>
</tr>
<tr>
<td>ILTASS</td>
<td>International long-term average speech spectrum</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>KEMAR</td>
<td>Knowles Electronics Manikin for Acoustic Research</td>
</tr>
<tr>
<td>MA</td>
<td>Maximum attenuation</td>
</tr>
<tr>
<td>MAP</td>
<td>Maximum a posteriori</td>
</tr>
<tr>
<td>MCL</td>
<td>Most comfortable level</td>
</tr>
<tr>
<td>MCRoomSim</td>
<td>Room simulator software</td>
</tr>
<tr>
<td>MLS</td>
<td>Maximum length sequence</td>
</tr>
<tr>
<td>MUSHRA</td>
<td>Multiple stimulus with hidden reference and anchor, sound quality test</td>
</tr>
<tr>
<td>NAL</td>
<td>National acoustic laboratories</td>
</tr>
<tr>
<td>NH</td>
<td>Normal hearing</td>
</tr>
<tr>
<td>NR</td>
<td>Cochlear Nucleus 6 SNR-NR algorithm</td>
</tr>
<tr>
<td>RM-ANOVA</td>
<td>Repeated measures analysis of variance</td>
</tr>
<tr>
<td>RT60</td>
<td>Room reverberation time</td>
</tr>
<tr>
<td>S</td>
<td>Standard directionality algorithm</td>
</tr>
<tr>
<td>SON90</td>
<td>Speech at 0°, noise at 90°</td>
</tr>
<tr>
<td>SCAN</td>
<td>Cochlear Nucleus 6 automatic classifier algorithm</td>
</tr>
<tr>
<td>SD</td>
<td>Standard deviation</td>
</tr>
<tr>
<td>SE</td>
<td>Standard error</td>
</tr>
<tr>
<td>SNR</td>
<td>Signal-to-noise ratio</td>
</tr>
<tr>
<td>SpatialNR</td>
<td>Spatial noise reduction algorithm</td>
</tr>
<tr>
<td>SRT</td>
<td>Speech reception threshold</td>
</tr>
<tr>
<td>SSQ</td>
<td>Speech, spatial and qualities of hearing scale</td>
</tr>
<tr>
<td>STI</td>
<td>Speech transmission index</td>
</tr>
<tr>
<td>SWN</td>
<td>Speech weighted noise</td>
</tr>
<tr>
<td>T30</td>
<td>Decay time estimation of reverberation time</td>
</tr>
<tr>
<td>TF</td>
<td>Time-frequency unit</td>
</tr>
<tr>
<td>T-SPL</td>
<td>Threshold sound pressure level</td>
</tr>
<tr>
<td>xPC</td>
<td>Mathworks Simulink xPC product</td>
</tr>
<tr>
<td>yrs</td>
<td>Years</td>
</tr>
<tr>
<td>Z</td>
<td>Zoom directionality algorithm</td>
</tr>
</tbody>
</table>
Mathematical symbols

\[ \hat{X}_k[n] \] Estimate of clean signal
\[ H_k[n] \] Noise reduction filter gain
\[ N_k[n] \] Noise reference signal
\[ S_k[n] \] Speech reference signal
\[ \xi_k[n] \] Instantaneous SNR
\[ k \] Frequency index
\[ n \] Time index
\[ \beta_A \] SpatialNR algorithm attack time parameter
\[ \beta_R \] SpatialNR algorithm release time parameter
\[ \alpha \] SpatialNR algorithm bias parameter
\[ \gamma \] SpatialNR algorithm maximum attenuation parameter
\[ p[n,k] \] Speech presence matrix as a function of time (n) and frequency (k)
\[ X_{CL}[n,k], \] Clean signal at the stimulus output measured in current levels as a function of time (n) and frequency (k)
\[ \hat{X}_{CL}[n,k] \] Estimate of clean signal at the stimulus output measured in current levels as a function of time (n) and frequency (k)
\[ E_{CL}[n,k] \] Signal error at the stimulus output measured in current levels as a function of time (n) and frequency (k)
\[ E_{dB}[n,k] \] Signal error at the stimulus output measured in dB SPL as a function of time (n) and frequency (k)
\[ e[n,k] \] Total errors as a function of time (n) and frequency (k)
\[ e_I[n,k], \] Type I errors as a function of time (n) and frequency (k)
\[ e_{II}[n,k], \] Type II errors as a function of time (n) and frequency (k)
\[ E_I \] Type I error rate over all time and frequency
\[ E_{II} \] Type II error rate over all time and frequency
\[ P(p,s) \] Proportion of votes as a function of program (p) and sound class (s)
\[ V(p,s) \] Number of votes as a function of program (p) and sound class (s)
\[ N_e(s) \] Normalised vote score for experimental program e as a function of sound class (s)
# Table of contents

Abstract ....................................................................................................................... iii

Declaration .................................................................................................................. v

Preface ...................................................................................................................... vii

Acknowledgments...................................................................................................... ix

Glossary ..................................................................................................................... xi
  Acronyms and abbreviations ................................................................................. xi
  Mathematical symbols ............................................................................................ xiii

Table of contents....................................................................................................... xv
  List of figures ........................................................................................................... xxii
  List of tables ............................................................................................................ xxvii

Chapter 1  Introduction ............................................................................................... 1
  1.1  CI device ........................................................................................................... 3
  1.2  Evaluating the effectiveness of noise reduction processing ......................... 4
    1.2.1  Theoretical measures ............................................................................... 4
           Speech transmission index (STI) ................................................................. 4
           Binary mask-based metric ......................................................................... 5
    1.2.2  Perceptual measures ............................................................................... 6
           Speech reception threshold (SRT) ............................................................... 6
           Sound quality (MUSHRA) ........................................................................... 7
           Acceptable noise level (ANL) .................................................................... 8
    1.2.3  Acoustic test environments ..................................................................... 9
    1.2.4  Field-based assessment ............................................................................ 11
# Table of contents

1.3 Noise reduction for CIs ................................................................. 11
  1.3.1 Single microphone noise reduction .................................. 11
  1.3.2 The binary mask ............................................................... 12
  1.3.3 Multi-microphone noise reduction ................................. 13
  1.3.4 Post-filter algorithms ......................................................... 14
1.4 Practical considerations ............................................................. 17
  1.4.1 Reverberation ................................................................. 17
  1.4.2 Microphone mismatch ....................................................... 19
1.5 Outline of the thesis ................................................................. 20
  1.5.1 Main research objectives .................................................. 20
  1.5.2 Chapter by chapter overview ............................................ 21

## Chapter 2 Algorithm Design .......................................................... 23
2.1 Introduction .................................................................................. 23
2.2 Algorithm kernel ......................................................................... 24
2.3 Theoretical analysis ..................................................................... 28
2.4 Algorithm parameters .................................................................. 29
  2.4.1 Signal smoothing ............................................................. 29
  2.4.2 Bias parameter ................................................................. 30
  2.4.3 Maximum attenuation ...................................................... 31
2.5 Summary ....................................................................................... 32

## Chapter 3 Methods ........................................................................ 33
3.1 Introduction .................................................................................. 33
3.2 Theoretical performance measures ........................................... 34
  3.2.1 STI.................................................................................... 34
  3.2.2 Error rate calculation ......................................................... 35
3.3 Laboratory-based perceptual measures ................................. 36
  3.3.1 Word recognition in quiet (CNC) ................................... 37
    CNC word test protocol ......................................................... 37
  3.3.2 Sentence recognition in noise (SRT) ............................... 37
    SRT test reliability ............................................................... 37
    SRT test protocol ................................................................. 39
  3.3.3 Sound quality rating (MUSHRA) .................................... 40
    MUSHRA test protocol .......................................................... 41
3.3.4 Acceptable noise level (ANL) ................................................................. 41
3.4 Noise Environment .................................................................................... 43
  3.4.1 Noise sources .......................................................................................... 43
  3.4.2 Spatial configuration ............................................................................... 44
    Summary of 2-SWN-rear configuration ....................................................... 46
    Summary of 4-TB-rear configuration ............................................................. 47
    Summary of 4-TB-full configuration ............................................................... 48
    Summary of 20-TB-full configuration ............................................................. 49
  3.4.3 Sound level and calibration ..................................................................... 50
  3.4.4 Stimulus presentation ............................................................................. 50
3.5 Test rooms ................................................................................................... 53
  3.5.1 Reverberant properties of the rooms ...................................................... 53
  3.5.2 Room simulator ...................................................................................... 54
3.6 Field-based outcome measures ................................................................... 54
  3.6.1 Questionnaire ......................................................................................... 54
  3.6.2 Voting ..................................................................................................... 55
    CP900 sound classification .......................................................................... 56
    CR230 remote control ............................................................................... 56
    Voting paradigm ........................................................................................... 56
    Voting data analysis .................................................................................... 58
3.7 Sound processing equipment ........................................................................ 59
  3.7.1 Simulink and xPC .................................................................................. 59
  3.7.2 CP900 sound processor ......................................................................... 59
  3.7.3 Slow and fast switching ......................................................................... 59
3.8 Subjects ........................................................................................................ 60
3.9 Experiment summary by outcome measure and noise type ....................... 61

Chapter 4 Experiment I Bias value in low reverberation ............................... 65
  4.1 Introduction ............................................................................................... 65
  4.2 Method ...................................................................................................... 65
    4.2.1 Processing conditions ......................................................................... 66
    4.2.2 Acoustic conditions .......................................................................... 66
    4.2.3 Outcome measures ............................................................................ 68
    4.2.4 Test protocol ....................................................................................... 68
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.3</td>
<td>Results</td>
<td>68</td>
</tr>
<tr>
<td>4.4</td>
<td>Discussion</td>
<td>70</td>
</tr>
<tr>
<td>4.4.1</td>
<td>SpatialNR parameter settings</td>
<td>70</td>
</tr>
<tr>
<td>4.4.2</td>
<td>Noise type and test SNR</td>
<td>72</td>
</tr>
<tr>
<td>4.4.3</td>
<td>Reverberation</td>
<td>72</td>
</tr>
<tr>
<td>4.4.4</td>
<td>Acclimatisation</td>
<td>72</td>
</tr>
<tr>
<td>4.5</td>
<td>Conclusion</td>
<td>73</td>
</tr>
</tbody>
</table>

**Chapter 5**  
**Experiment II**  
Bias value and time constants in reverberation | 75 |
| 5.1 | Introduction | 75 |
| 5.2 | Algorithm parameter selection | 76 |
| 5.3 | Method | 76 |
| 5.3.1 | Processing conditions | 79 |
| 5.3.2 | Acoustic conditions | 80 |
| 5.3.3 | Outcome measures | 80 |
| 5.3.4 | Test protocol | 80 |
| 5.4 | Results | 81 |
| 5.4.1 | SRT | 81 |
| 5.4.2 | MUSHRA | 83 |
| 5.4.3 | CNCs | 85 |
| 5.5 | Discussion | 85 |
| 5.5.1 | SpatialNR | 85 |
| 5.5.2 | Zoom/Beam | 86 |
| 5.5.3 | NR | 87 |
| 5.5.4 | CI performance in reverberation | 87 |
| 5.5.5 | Future research | 87 |
| 5.6 | Conclusion | 87 |

**Chapter 6**  
**Experiment III**  
High reverberation | 89 |
| 6.1 | Introduction | 89 |
| 6.2 | Methods | 89 |
| 6.2.1 | Room simulator | 90 |
| 6.2.2 | Processing conditions | 92 |
6.2.3 Acoustic conditions ................................................................. 92
6.2.4 Outcome measures .............................................................. 93
6.2.5 Test protocol ......................................................................... 93
        Protocol deviations ............................................................. 93
6.3 Results ..................................................................................... 93
    6.3.1 Physical distance to sound sources .................................. 94
    6.3.2 Room simulator validation .............................................. 95
    6.3.3 Very high reverberation .................................................. 96
6.4 Discussion .............................................................................. 98
    6.4.1 Physical distance to sound sources .................................. 98
    6.4.2 Room simulator validation .............................................. 99
    6.4.3 Very high reverberation .................................................. 99
6.5 Conclusion .............................................................................. 100

Chapter 7  Experiment IV  Microphone mismatch ............................... 101

Chapter 8  Experiment V  BTE device take-home evaluation ............... 111
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Implementation platform</td>
<td>113</td>
</tr>
<tr>
<td>8.2.2 Laboratory-based evaluation</td>
<td>114</td>
</tr>
<tr>
<td>8.2.3 Field-based evaluation</td>
<td>114</td>
</tr>
<tr>
<td>Zoom-SpatialNR(MA14)</td>
<td>116</td>
</tr>
<tr>
<td>SCAN-SpatialNR(MA14)</td>
<td>116</td>
</tr>
<tr>
<td>SCAN-SpatialNR(MA7)</td>
<td>116</td>
</tr>
<tr>
<td>SCAN-NR</td>
<td>117</td>
</tr>
<tr>
<td>Directionality</td>
<td>118</td>
</tr>
<tr>
<td>8.2.4 Test protocol</td>
<td>118</td>
</tr>
<tr>
<td>8.3 Laboratory-based results</td>
<td>119</td>
</tr>
<tr>
<td>8.3.1 Overall performance</td>
<td>119</td>
</tr>
<tr>
<td>SRT</td>
<td>119</td>
</tr>
<tr>
<td>MUSHRA</td>
<td>121</td>
</tr>
<tr>
<td>ANL</td>
<td>121</td>
</tr>
<tr>
<td>8.3.2 Effect of MA</td>
<td>121</td>
</tr>
<tr>
<td>SRT</td>
<td>122</td>
</tr>
<tr>
<td>MUSHRA</td>
<td>123</td>
</tr>
<tr>
<td>ANL</td>
<td>123</td>
</tr>
<tr>
<td>8.3.3 Effect of SNR-NR</td>
<td>123</td>
</tr>
<tr>
<td>8.3.4 Effect of acclimatisation</td>
<td>125</td>
</tr>
<tr>
<td>8.3.5 Effect of implementation platform</td>
<td>125</td>
</tr>
<tr>
<td>8.3.6 Normal hearing benchmark</td>
<td>125</td>
</tr>
<tr>
<td>8.4 Field-based results</td>
<td>128</td>
</tr>
<tr>
<td>8.4.1 Individual voting preference</td>
<td>128</td>
</tr>
<tr>
<td>8.4.2 Group voting preference</td>
<td>133</td>
</tr>
<tr>
<td>8.4.3 Questionnaire</td>
<td>134</td>
</tr>
<tr>
<td>SSQ results</td>
<td>134</td>
</tr>
<tr>
<td>Overall preference results</td>
<td>134</td>
</tr>
<tr>
<td>8.5 Discussion</td>
<td>136</td>
</tr>
<tr>
<td>8.5.1 Overall laboratory-based performance</td>
<td>136</td>
</tr>
<tr>
<td>8.5.2 Effect of MA</td>
<td>137</td>
</tr>
<tr>
<td>8.5.3 Effect of SNR-NR</td>
<td>137</td>
</tr>
<tr>
<td>8.5.4 Effect of acclimatisation</td>
<td>137</td>
</tr>
<tr>
<td>8.5.5 Effect of implementation platform</td>
<td>138</td>
</tr>
</tbody>
</table>
List of figures

Figure 1-1: Typical speech at 0 degrees with co-located 4 talker babble noise source at 90 degrees (S0N90) e.g. (Spriet et al., 2007)........................................................................................................10

Figure 1-2: Spatialized HINT test (Nilsson et al., 2005), each noise source is co-located 4 talker babble, 4x4=16 competing talkers in total........................................................................................................10

Figure 1-3: Top view illustrating the relative location of multiple microphones of a behind-the-ear (BTE) device. An acoustic path difference exists, which is dependent on the direction of arrival of sound........................................................................................................................................14

Figure 2-1: Signal path integration used to evaluate the SpatialINR algorithm. The speech reference (Standard or Zoom) and noise reference signals are used as inputs to Beam and SpatialINR algorithms. A switch provides selection of either Beam or SpatialINR........................................25

Figure 2-2: SpatialINR algorithm kernel. The two microphone signals were used to create two spatially filtered signals; a forward-facing speech reference (SFB) and a rear-facing noise reference (NB). The SNR (\(\xi dB\)) was calculated from the ratio of the smoothened speech and noise reference signals, which was used to calculate the noise reduction gains using a parametric Wiener-like function. ........................................................................................................................................26

Figure 2-3: (A) Single source spatial sensitivity of Speech Reference (Zoom directionality) and Noise Reference signals and (B) the resulting single source SNR \(\xi dB\) according to Equation 2-3. Sensitivity was recorded with a broadband speech-shaped noise stimulus presented under anechoic conditions with device worn on the left ear of a KEMAR manikin. The relative sensitivity is the RMS value across the duration of the stimulus averaged across all frequency channels. ........................................................................................................................................27

Figure 2-4: SNR \(\xi dB\) as a function of instantaneous input SNR for target speech from 0° embedded in speech-shaped noise from six different locations ranging from 30° to 180°. \(\xi dB\) is upper-bounded by the single source SNR contribution of the target speech, and lower bounded by the single source SNR contribution of the noise which varies with the noise’s location.......28

Figure 2-5: Effect of signal smoothing parameter, \(\beta\), for a step input change in input signal level showing symmetric and asymmetric attack and release time constants. .................................29

Figure 2-6: Parametric Wiener filter which maps Input SNR, \(\xi\), to Gain. The effect of bias parameter, \(\alpha\), is shown for a range of values..................................................................................................................30

Figure 2-7: Single source spatial sensitivity in anechoic conditions of Standard, Zoom, Beam and SpatialINR processing including effect of SpatialINR bias parameter, \(\alpha\).................................................31

Figure 2-8: Gain function (\(\alpha = 0\ dB\)) with attenuation that is limited by the maximum attenuation parameter, \(y\)........................................................................................................................................32

Figure 3-1: Example screen shot from the SRT test. The left of this screen shows test configuration settings and the right hand plot shows the adaptation of an SRT run where the SNR adapted up and down during the test based on the subject’s responses.................................................................38

Figure 3-2: Test reliability in randomly allocated spatially separated 4-talker babble (4-TB-rear).............38

Figure 3-3: Example subject input screen used in the MUSHRA sound quality evaluation. The subject was asked to listen to the stimulus processed with each of the different conditions by pressing the buttons labelled ‘Reference’ and ‘A’-‘F’ then rate the overall sound quality of each sample according to the scale from ‘Very Poor’ to ‘Excellent’..............................42

Figure 3-4: User interface used to conduct the acceptable noise level (ANL) test. This screen shows the up/down buttons used to change the level of running speech without background noise present in order to obtain the most comfortable level (MCL)........................................................................42
Figure 3-5: User interface used to conduct the acceptable noise level (ANL) test. This screen shows the up/down buttons used to change the level of the noise to obtain the acceptable noise level. ...........................................................................................................................................42
Figure 3-6: Loudspeaker locations used for 2-SWN-rear noise configuration .........................................................46
Figure 3-7: Loudspeaker locations used for 4-TB-rear noise configuration .................................................................47
Figure 3-8: Loudspeaker locations used for 4-TB-full noise configuration .................................................................48
Figure 3-9: Loudspeaker locations used for 20-TB-full noise configuration ...............................................................49
Figure 3-10: The sound treated Whitford booth located at Cochlear, East Melbourne, used for speech tests and evaluation. Reverberation time $T_{30} = 87$ ms..........................................................................................52
Figure 3-11: Reverberant training room used for evaluation of the noise reduction algorithms. Reverberation time $T_{30} = 523$ ms..................................................................................................................52
Figure 3-12: Example response sheet from the modified SSQ questionnaire that allowed simultaneous ratings of two listening programs, not possible with the original SSQ format. ..............................55
Figure 3-13: The CP900 sound processor and handheld CR230 remote control used in the voting paradigm ...........................................................................................................................................57
Figure 4-1: Theoretical performance of SpatialNR algorithm with different bias value and input signal directionality, measured in 4-TB-rear noise in the sound treated Whitford booth. Increasing the bias value provides more aggressive noise reduction. .........................................................67
Figure 4-2: Individual SRT results in spatially separated 4-talker babble from the rear hemi-field for Beam and each of the five SpatialNR conditions. Lower scores indicate improved performance..................................................................................................................68
Figure 4-3: SRT benefit relative to Beam in spatially separated 4-talker babble from the rear hemi-field. Box plots show the 10, 25, 50, 75 and 90th percentiles and all outliers are marked with dots. The means are shown as dotted lines. Stars indicate statistically significance ($P<0.001$) ....69
Figure 5-1: Theoretical performance of SpatialNR (smoothing 5/50 ms) with different bias value. Calculated in 4-TB-rear and 20-TB-full noise at 0 dB SNR in the reverberant training room.77
Figure 5-2: Theoretical performance of SpatialNR for different smoothing parameters (using fixed bias value +3 dB) in 4-TB-rear (left panels) and 20-TB-full (right panels) noise types. Heat maps shows type I errors (upper panels) and type II errors (lower panels) with higher error rates shaded darker and contour intervals of 1%. For reference, open and closed points indicate the 10/10 and 5/50 ms combinations used in clinical testing. Acoustic stimulus was presented in the NAL Training room using running speech in noise at 0 dB SNR.................78
Figure 5-3: Group mean SRTs in 4-talker rear half (4-TB-rear) and 20-talker full circle (20-TB-full) noise for all processing conditions evaluated in Experiment II. Lower scores indicate better performance. Error bars show the 95% confidence interval .................................................................82
Figure 5-4: SRT benefit over Beam baseline for different SpatialNR bias and signal smoothing parameters. SRT benefit was calculated as Beam SRT minus SpatialNR SRT. Error bars indicate 95% confidence intervals .................................................................83
Figure 5-5: Group mean sound quality ratings for the MUSHRA task in 4-talker rear half (4-TB-rear) and 20-talker full circle (20-TB-full) noise at two different input SNRs. Error bars indicate the 95% confidence interval. SpatialNRa and SpatialNRd were omitted from evaluation in the MUSHRA task. The Reference condition used Standard processing at an input SNR of +20 dB. Data from Beam and BeamNR conditions were not able to be interpreted due to lack of adaptation of the Beam algorithm within the test framework.........................................................84
Figure 5-6: Group mean CNC word scores in quiet. Error bars indicate 95% confidence intervals. SpatialNRa and SpatialNRd were omitted from CNC word testing. .................................................................85
Figure 6-1: The energy decay curves comparing the measured and simulated impulse responses from the front left microphone at a distance of (A) 1, (B) 2 and (C) 3 m.................................91
Table of contents

Figure 6-2: SRT results for each processing condition in (A) Low R, (B) Mid R and (C) High R reverberation conditions. Error bars show the 95% C.I. ................................................................. 94

Figure 6-3: SRT Benefit over the Omni processing conditions. Data has been averaged across three reverberation conditions Low R, Mid R and High R. Error bars show the 95% C.I. .................................................. 94

Figure 6-4: SRT Results for (A) High R and (B) HighSim reverberation conditions. Error bars indicate 95% C.I. ............................................................................................................................................ 96

Figure 6-5: SRT results for HighSim and ExtraHighSim reverberation conditions. Error bars indication the 95% C.I. Note the change of scale from previous figures with similar data. ................................. 96

Figure 6-6: SRT benefit of each processing condition over the Omni conditions, in HighSim and ExtraHighSim levels of reverberation. Error bars indicate 95% C.I. ................................................................. 97

Figure 7-1: Broad-band spatial response of (A) Zoom and (B) noise reference with different levels of microphone mismatch indicated with shading. The plots show the directional sensitivity in dB under anechoic conditions when the microphones are worn on KEMAR manikin left ear. For each level of microphone mismatch, the sensitivity at each angle is relative to 0 degrees (Zoom) or 180 degrees (noise reference). .................................................................................. 103

Figure 7-2: (A) Type I error rate, (B) type II error rate and (C) speech transmission index (STI) for various levels of microphone mismatch. Noise type used was 4-TB-full at 0 dB SNR in the Whitford sound treated booth. .................................................................................................................................. 104

Figure 7-3: Speech intelligibility in 4-TB-full as a function of microphone sensitivity mismatch. Group mean SRTs are shown and error bars show the 95% confidence interval ........................................... 107

Figure 7-4: MUSHRA sound quality ratings in 4-TB-full noise for various levels of microphone mismatch. Group mean data are shown with error bars indicating the 95% confidence interval ...... 107

Figure 8-1: Group SRT, MUSHRA and ANL results for BTE implementation of the four main directional algorithms (Standard, Zoom, Beam(Z) and SpatialNR(MA14)) that were evaluated, separated by noise type. Error bars show 95% confidence interval. Post-hoc comparisons between processing conditions that were statistically significant are marked with asterisks (*P<0.05, **P<0.01 and ***P<0.001). ....................................................................................................................................... 120

Figure 8-2: The benefit of maximum attenuation on speech intelligibility (SRT), sound quality (MUSHRA) and acceptable noise level (ANL) of single channel noise reduction (NR) evaluated in combination with SpatialNR(MA14) in each of the four noise types. The benefit of NR on relative to NR off is shown, error bars show 95% confidence intervals. Post-hoc comparisons that were statistically significant are marked with asterisks (*P<0.05, **P<0.01 and ***P<0.001). ....................................................................................................................................... 122

Figure 8-3: Benefit in speech intelligibility (SRT), sound quality (MUSHRA) and acceptable noise level (ANL) of single channel noise reduction (NR) evaluated in combination with SpatialNR(MA14) in each of the four noise types. The benefit of NR on relative to NR off is shown, error bars show 95% confidence intervals. Post-hoc comparisons that were statistically significant are marked with asterisks (*P<0.05, **P<0.01 and ***P<0.001). ....................................................................................................................................... 124

Figure 8-4: The effect of exposure to Zoom and SpatialNR(MA14) on speech intelligibility outcomes. SRTs in 4-TB-rear noise are presented with error bars showing 95% confidence intervals. 126

Figure 8-5: The effect of exposure to SpatialNR on speech intelligibility outcomes presented as the difference between Zoom and SpatialNR(MA14). Error bars show 95% confidence intervals. ....................................................................................................................................... 126

Figure 8-6: Effect of implementation platform on speech intelligibility. SRT’s are shown for four pairs of conditions (connected with a line) that were identical apart from the implementation platform. Error bars show 95% confidence intervals ....................................................................................................................................... 127

Figure 8-7: Unaided normal hearing (NH) SRTs in comparison to CI performance using both Standard directionality and with SpatialNR(MA14). NH were evaluated in three of the four different noise types. Data points show the mean and 95% confidence intervals .................................................. 127
Figure 8-8: Individual raw voting data for comparisons (A) Zoom-SpatialNR(MA14), (B) SCAN-SpatialNR(MA14), (C) SCAN-SpatialNR(MA7) and (D) SCAN-NR. Size of bubble indicates the number of votes while colour indicates the sound class in which votes were made. Data are plotted on the x-axis as the normalised vote score such that the zero line indicates equal votes in each program. .................................................................................................................................................................................. 129

Figure 8-9: Distribution of votes for the directionality comparison. The location of the data points indicate the proportion of votes in each of the three programs, where the tips of the triangle plot indicate all votes for that program. The size of the bubble indicates the number of votes and the colour indicates the sound class. ......................................................................................................................... 130

Figure 8-10: Individual vote scores for the Zoom-SpatialNR(MA14), SCAN-SpatialNR(MA14) and SCAN-NR comparisons. Plots show association of vote scores across different sound classes. Solid lines represent the model $y=ax$. Dotted lines show the 95% C.I. for the estimate of the model parameter, $a$. A positive gradient on both the upper and lower 95% C.I. indicates the model fit was significant. .................................................................................................................................................................................. 131

Figure 8-11: Group summary of voting data for the first four voting comparisons (A) Zoom-SpatialNR(MA14), (B) SCAN-SpatialNR(MA14), (C) SCAN-SpatialNR(MA7) and (D) SCAN-NR. Median vote scores for the three dominant sound classes are shown with p-values for one-sample signed rank tests against a hypothesised median score of 0 (*$p<0.05$, **$p<0.01$ and ***$p<0.001$). The table indicates the number of subjects that contribute to the median scores, divided by overall preference into categories against/equal for SpatialNR/SNR-NR. The mean number of vote events per subject and scene are also shown.................................................. 132

Figure 8-12: Distribution of sound class independent vote score for the first four voting comparisons. Box and whisker plots show the 10, 25, 50, 75, and 90th percentiles. The mean line is marked with a dotted line and all outliers are shown as dots. ................................................................. 133

Figure 8-13: Questionnaire-based overall preference scores in quiet and noisy conditions for (A) Zoom-SpatialNR(MA14), (B) SCAN-SpatialNR(MA14), (C) SCAN-SpatialNR(MA7) and (D) SCAN-NR comparisons. Negative score indicates preference for baseline condition and positive score indicates preference for the experimental condition. Degree of preference is given by the magnitude of the score: |3|="Much prefer", |2|="Moderately prefer", |1|="Slightly prefer", 0="No difference"................................................................. 135

Figure 8-14: Group summary of questionnaire-based overall preference score for quiet and noisy conditions for each of the four comparisons. Box and whisker plots indicate the 10, 25, 50, 75 and 90th percentiles with dashed mean line and all outliers marked as dots. Asterisks indicate group median score significantly different from hypothesised median of zero (*$p<0.05$)........................................................................................................................................................................................................... 136

Figure 8-15: Number of vote events vs. number of listening hours, separated by take-home comparison on separate panels. The sound class is identified by the marker colour, blue=Quiet, green=Speech in noise and orange=Noise. Markers show the median and error bars show the first and third quartiles ........................................................................................................................................................................................................... 142

Figure 8-16: Rate of voting (votes/hr) compared with rate of changing programs (changes/hr) for each of the voting comparisons on separate panels. The sound class is identified by the marker colour, blue=Quiet, green=Speech in noise and orange=Noise. Markers show the median and error bars show the first and third quartiles ........................................................................................................................................................................................................... 143

Figure 8-17: Relationship between voting preference and listening time for the Zoom-SpatialNR(MA14), SCAN-SpatialNR(MA14) and SCAN-NR comparisons. The normalised proportion of listening time is plotted against the normalised vote score for each sound class. Bubble size indicates number of votes for each data point (<3 votes not shown). Colour indicates sound class (blue=Quiet, green=Speech in Noise, orange=Noise). Solid lines represent the model $y=ax$. 
Dotted lines show the 95% C.I. for the estimate of the model parameter, $a$. A positive gradient on both the upper and lower 95% C.I. indicates the model fit was significant. ....146

Figure 8-18: Relationship between laboratory and field based assessment for the Zoom-SpatialNR(MA14) comparison. SRT, MUSHRA and ANL benefit was calculated as the difference between the SpatialNR(MA14) and Zoom conditions. Normalised vote scores were summed across all sound classes. Solid lines represent the least squares regression to the model $y = ax$. Dotted lines show the 95% C.I. for the estimate of the model parameter, $a$. A positive gradient on both the upper and lower 95% C.I. indicates the model fit was significant. ..........................................................147

Figure 8-19: Relationship between laboratory and field-based assessment for the Quiet sound class of the SCAN-SpatialNR(MA14) comparison. SRT, MUSHRA and ANL benefit was calculated as the difference between the SpatialNR(MA14) and Standard conditions. Normalised vote scores were from the Quiet sound class. Solid lines represent the least squares regression to the model $y = ax$. Dotted lines show the 95% C.I. for the estimate of the model parameter, $a$. A positive gradient on both the upper and lower 95% C.I. indicates the model fit was significant. ..........................................................148

Figure 8-20: Relationship between laboratory and field-based assessment for the Noise sound class of the SCAN-SpatialNR(MA14) comparison. SRT, MUSHRA and ANL benefit was calculated as the difference between the SpatialNR(MA14) and Zoom conditions. Normalised vote scores were from the Noise sound class. Solid lines represent the least squares regression to the model $y = ax$. Dotted lines show the 95% C.I. for the estimate of the model parameter, $a$. A positive gradient on both the upper and lower 95% C.I. indicates the model fit was significant. ..........................................................149

Figure 8-21: Relationship between laboratory and field-based assessment for the Speech in Noise sound class of the SCAN-SpatialNR(MA14) comparison. SRT, MUSHRA and ANL benefit was calculated as the difference between the SpatialNR(MA14) and Beam(Z) conditions. Normalised vote scores were from the Speech in Noise sound class. Solid lines represent the least squares regression to the model $y = ax$. Dotted lines show the 95% C.I. for the estimate of the model parameter, $a$. A positive gradient on both the upper and lower 95% C.I. indicates the model fit was significant. ..........................................................150

Figure 8-22: Relationship between laboratory and field-based assessment for the SCAN-NR comparison. SRT, MUSHRA and ANL benefit was calculated as the difference between the SpatialNR(MA14) and SpatialNR(MA14) conditions. Normalised vote scores were summed across all sound classes. Solid lines represent the least squares regression to the model $y = ax$. Dotted lines show the 95% C.I. for the estimate of the model parameter, $a$. A positive gradient on both the upper and lower 95% C.I. indicates the model fit was significant. ..........................................................150

Figure 8-23: Correlation of overall preference score and normalised vote score for the Zoom-SpatialNR(MA14), SCAN-SpatialNR(MA14) and SCAN-NR comparisons. Both axes indicate degree of preference where negative scores indicate a preference for SCAN and positive scores a preference for SpatialNR(MA14) or NR. Size of bubble indicates number of votes recorded for that data point (<3 votes not shown). Colour of bubble indicates the listening scene (blue=Quiet, orange=Noise). Solid lines represent the model $y=ax$. Dotted lines show the 95% C.I. for the estimate of the model parameter, $a$. A positive gradient on both the upper and lower 95% C.I. indicates the model fit was significant. ..........................................................152
List of tables

Table 3-1: Test reliability reported as the standard deviation (SD) in dB between test re-test SRT for various speech in noise tests, for normal hearing (NH) and cochlear implant (CI) subjects...39
Table 3-2: Spatial configurations for 2-SWN-rear noise type showing number of noise sources at each spatial location for the different tests. (* indicates randomly chosen locations)....................46
Table 3-3: Spatial configurations for 4-TB-rear noise type showing number of noise sources at each spatial location for the different tests. (* indicates randomly chosen locations)....................47
Table 3-4: Spatial configurations for 4-TB-full noise type showing number of noise sources at each spatial location for the different tests. (* indicates randomly chosen locations)....................48
Table 3-5: Spatial configurations for 20-TB-full noise type showing number of noise sources at each spatial location for the different tests. (* indicates randomly chosen locations)....................49
Table 3-6: Room reverberation properties for the rooms and listening distances used in the evaluation. Reverberation time (T30), early decay time (EDT), clarity indices (C50, C7) and direct-to-reverberant ratio (DRR) have been estimated from the impulse response data at each listening distance.................................................................54
Table 3-7: CP900 SCAN sound classes and associated directionality that is automatically selected......56
Table 3-8: Summary of outcome measures and noise types used in each of the five experiments.......61
Table 3-9: Demographic details and cochlear implant map configuration for the subjects who participated in the experiments. R – right ear, L – left ear, HA – hearing aid .........................62
Table 4-1: Configuration of SpatialNR processing conditions evaluated in Experiment I................66
Table 5-1: Configuration of SpatialNR processing conditions evaluated in Experiment II............79
Table 5-2: Summary of processing conditions and outcome measures used in the evaluation during Experiment II.................................................................79
Table 6-1: Acoustic properties of the training room, calculated from the measured and simulated impulse responses. The data presented were calculated as the median over each set of 12 values (calculated at angles from 0 to 330 degrees, in 30 degree increments)...................92
Table 6-2: Configuration of SpatialNR processing condition evaluated in Experiment III ...............92
Table 6-3: Physical separation of target and noise source from the listener and method of impulse response generation for each of the five reverberant configurations used for evaluation in Experiment III.................................................................93
Table 7-1: Configuration of SpatialNR processing condition evaluated in Experiment IV................103
Table 7-2: Processing conditions that were compared simultaneously for each instance of the MUSHRA sound quality task.................................................................106
Table 8-1: Configuration of SpatialNR processing conditions evaluated in laboratory as part of Experiment V.................................................................113
Table 8-2: Processing conditions used for evaluation, showing details of sound processing, implementation platform and laboratory-based tests that were involved. ..115
Table 8-3: Configuration of microphone directionality for modified SCAN programs used for field-based assessment. SCAN was the CP900 default setting not used in this experiment, SCAN1 was used as the base program (SpatialNR disabled), and SCAN2 was used as the test program with SpatialNR enabled. .................................................................116
Table 8-4: Program configurations used for the field-based assessment. With the exception of the ‘Directionality’ comparison, the underlying directionality in each sound class was consistent across the two programs being evaluated, such that only the experimental algorithm varied. ..................................................................................................117
Chapter 1

Introduction

Cochlear implant (CI) users can achieve good performance in quiet, low noise conditions, but when conditions become noisy, speech understanding is compromised. CI users’ ability to communicate in noisy situations is limited and improving performance in noise remains a challenging area of CI research.

In order to improve speech understanding in noisy conditions, signal processing is used to remove some or all of the competing noise, whilst maintaining the speech target signal with little or no modification. To achieve this, noise reduction algorithms utilise either single microphone or multiple microphones inputs. Single microphone techniques rely on assumptions of the acoustic features related specifically to speech and noise, predominantly the fact that the level of speech signals change more frequently and more quickly than that of noisy signals. The noise reduction processing can use these differences to try and separate speech from noise. Multiple microphone algorithms have an added advantage in that they can use information about the direction of arrival of sounds, and can filter the signal based on spatial location. Well-known multi-microphone techniques use directional microphones, also known as beamforming, to selectively reduce the level of signals arriving from behind and beside the listener.

Both single microphone and multi-microphone noise reduction techniques have been evaluated in CI listeners, demonstrating that CI performance can be improved under controlled acoustic conditions. The precise nature of the acoustic environment used for evaluation can have a large impact on the outcome and the effectiveness of noise reduction algorithms over a wide range of acoustic conditions. Effectiveness is typically limited to certain noise types, spatial configurations and acoustic environments. Single microphone noise reduction can provide speech intelligibility improvement when the background noise is unmodulated, steady-state noise (e.g. speech shaped Gaussian random noise). However, when the background noise is modulated, such as when the noise is comprised of competing talkers, the effectiveness of single microphone noise reduction is limited. Multi-microphone noise reduction can provide the most speech intelligibility improvement when the background noise originates from a single location separated from the target speech, but real-life situations don’t often meet this ideal. When the number of competing sources is increased or the
acoustics of the room are highly reverberant, the effectiveness of multi-microphone noise reduction declines. Performance is also compromised when the microphones in the array are not well matched in sensitivity and phase to each other.

In the study of single microphone noise reduction in CIs, the ideal binary mask has provided interesting insights in experiments when the speech and noise signals are known a priori. In these studies, prior knowledge of the ideal binary mask, that is, prior knowledge of the signal-to-noise ratio (SNR) at each frequency and point in time, allows recovery of the original speech signal. This illustrates that if the SNR is accurately known, the deleterious effects of noise can be removed. In CI noise reduction systems, this knowledge is not directly available, and the SNR must be estimated from the microphone input signal. In this project, the task of estimating the SNR is based on multiple microphones, whereas traditionally this estimation has fallen under the domain of single microphone techniques.

The main focus of this project is to present and evaluate a noise reduction system for CI that is based on estimating the SNR from multiple microphones. Certain aspects of single microphone and multi-microphone noise reduction systems are combined for the purpose of improving performance for CI users in a wide variety of noisy conditions. This overcomes limitations of existing, separate techniques. In chapter 2, a noise reduction system is proposed, where the SNR is estimated using spatial information contained in multiple microphones. The SNR is then used to produce a cleaned output signal by reducing the level of noisy channels. Unlike single microphone systems that estimate the SNR based on assumptions of signal statistics, the proposed algorithm produces an estimate of the SNR that is based on direction. The system differs from multi-microphone linear filtering techniques (e.g. beamforming, multi-channel Wiener filter), by introducing non-linear operations used in single microphone noise reduction. This results in a hybrid architecture involving aspects of both single and multi-microphone approaches.

The thesis essentially documents the evolution of the algorithm, beginning with experiments using a laboratory-based real-time system and concluding with a behind-the-ear (BTE) speech processor that allowed evaluation of the algorithm by CI users away from the laboratory. The algorithm is evaluated in a range of noisy conditions, with the aim of understanding the algorithm’s performance in environments that represent difficult situations for CI users. The algorithm is evaluated with CIs listeners primarily using a speech intelligibility task in controlled acoustic environments, while sound quality and acceptable noise level are also assessed. Existing commercial algorithms from Cochlear Ltd are used as baseline conditions for comparison. The proposed noise reduction system has various tuneable parameters that are systematically varied under a range of different test conditions. This enables an understanding of the impact of each parameter on noise reduction performance. In chapters 4 and 5, the algorithm’s bias parameter is systematically adjusted and the effect on speech intelligibility is assessed while the test room was altered to include reverberation. In chapter 6, a more thorough examination is made of the effects of reverberation on noise reduction performance. The effect of microphone mismatch is assessed in chapter 7. Complementary to assessments made in the laboratory, chapter 8 explores the field-based response of CI users to the proposed algorithm. In particular, the field-based assessment allowed users to directly vote for
their preferred settings using their sound processor’s remote control. The enables users to provide direct feedback on their listening experience at the time and place of evaluation.

1.1 CI device

A cochlear implant (CI) is an implanted medical device used to restore hearing loss associated with inner hair cell depletion. CI systems consist of an externally worn sound processor and an internally implanted stimulator used to electrically excite the auditory nerve of an impaired ear. Electrical stimulation is provided by a multi-contact electrode array that is surgically placed in the cochlea. The implant’s electronics connected to the electrode array are contained in a hermetically sealed container, which is also implanted under the skin on the mastoid. Communication between the externally worn sound processor and the implant electronics is via a radio-frequency wireless link. A magnet is contained both in the implant and transmitting coil to hold the external coil in place during use.

The sound processor contains the power source, microphone(s), digital signal processor (DSP), and radio-frequency transmission system for transmitting data and power across the skin to the implant. The implant decodes the transmission and generates electrical impulses that are delivered via the multi-electrode array implanted in the cochlea.

The stimulation of nerve cells by the CI aims to replace the excitation that normally occurs via intact healthy inner hair cells. Traditional signal processing for CI devices usually involves extracting the frequency components of the input signal and assigning them to different electrodes in the stimulating array. The neurons residing near each electrode are selectively stimulated with the energy corresponding to a particular frequency, which would normally be stimulated if the cochlea was healthy. Typically, the amplitude envelope in different spectral channels is analysed and the amplitude of electrical stimulation is derived from the envelopes. The stimulation paradigm involves sequentially stimulating each electrode with the stimulation level derived from the acoustic amplitude. When all available electrodes on the array are stimulated in each stimulation cycle, the stimulation strategy is referred to as continuous interleaved sampling (CIS, Wilson et al., 1991). Another approach is the so-called n-of-m strategy, where only a subset of frequency channels is selected for stimulation in any particular cycle. The Advanced Combination Encoder (ACE) strategy uses n channels that have largest energy in the spectral envelope (McDermott et al., 1992; Vandali et al., 2000), with typical clinical default value of n=8, and usually in the range from 6 to 12. In this context, the n-of-m strategy is often termed maxima selection, referring to the process of selecting channels with the highest amplitude.

CI recipients often need to communicate in noisy situations in their everyday lives, where speech understanding is poorer than that usually obtained in quiet conditions. Speech intelligibility performance of CI users is well known to degrade in noisy conditions. Signal processing can be used to enhance the intelligibility of speech signals before they are delivered to the implant. Efforts have concentrated on using signal processing to reduce the disturbance of noise with the ultimate aim of improving speech intelligibility in noisy conditions. The
evaluation of noise reduction in CI speech processing is a key element to determine the effectiveness and overall benefit that can be provided.

1.2 Evaluating the effectiveness of noise reduction processing

In order to evaluate the effectiveness of noise reduction processing, the best test is to involve human listeners. Speech intelligibility and sound quality are typical outcomes that can be measured reliably under controlled acoustic conditions, making the data amenable to statistical analysis. Additionally, many researchers use theoretical measures that can be calculated without the involvement of human listeners. The drawback of this approach is the predictive power of the theoretical measures, which can be restricted to certain types of acoustic processing. This limits the generalisation of theoretical approaches in predicting outcomes for human listeners.

Regardless of whether human listeners are involved in the assessment, the specific acoustic situation that is used for the test has a large impact on how the results are to be interpreted. This is particularly pertinent in the case of CI devices because ultimately the device is to be used in a wide range of acoustic environments of users in their day-to-day lives. Therefore, both theoretical and laboratory-based experiments lack a direct connection to how the product is used day to day, and the acoustic situation used for evaluation needs to be carefully contrived.

To overcome this, field-based assessment can be used, which allows CI users to experience the experimental algorithm in their every-day lives. Typically, this listening experience is accompanied by a formal questionnaire targeting structured feedback from the listening experience. This type of subjective assessment can provide useful information to support the laboratory-based outcomes.

1.2.1 Theoretical measures

Before proceeding to clinical evaluation with human listeners, it is beneficial to have an objective metric that can be calculated to predict speech performance, even when the predictive power is not strong. Theoretical measurements of algorithm performance can provide insight into the operation of the algorithm, provide feedback to the developer about potential deficiencies or problem areas, and generally allow comparisons of different algorithm variations.

Speech transmission index (STI)

Two commonly used metrics that are based on weighted frequency band calculations of SNR are the Speech Transmission Index (STI, Steeneken and Houtgast, 1980) and the Articulation Index (AI, Fletcher and Galt, 1950), which has been adapted to form the Speech Intelligibility Index (SII, ANSI S3.5-1997). Both of these methods are based on signals processed by a ‘black box’ which requires analysis of acoustic or time domain outputs not necessarily available in CI
speech processing. Although these metrics predict speech intelligibility well for linear filtering and additive steady state noise, they do not predict perception well for non-linear processing such as noise reduction and automatic gain control. In addition, these metrics are designed to predict speech intelligibility for normal hearing or hearing impaired listeners, not for CI listeners.

Goldsworthy and Greenberg (2004) introduced four speech-based STI metrics with the aim of overcoming the inadequacies of traditional and other speech-based STI approaches. All proposed methods that were presented agreed with the original STI for additive noise and linear operations, although there was a non-linear relationship between the normalised correlation method and original STI. This means the absolute STI value according to the normalised correlation method might be difficult to interpret, but relative comparisons of STI values calculated within the metric are suitable. The normalised correlation method represents the biggest departure from traditional STI and provides an important alternative to traditional STI. In addition, the STI calculation procedures can be tailored to match a particular CI speech processor by matching the frequency bands and method of envelope calculation (Goldsworthy and Greenberg, 2004). These authors also suggested the method was appropriate for predicting the effect of spectral subtraction processing for CI users. The STI metric has been used to predict vocoded speech in noise and reverberation (Whitmal and Poissant, 2009) and to predict CI performance with fixed and adaptive directional microphones (Chung et al., 2006).

Binary mask-based metric

The binary mask concept involves decomposing the signal into a time-frequency (TF) representation and assigning a binary decision to each TF element depending on whether the element is dominated by speech or noise. The group of binary decisions is called a mask which can be applied to the signal and, with correct identification of speech and noise elements, used to remove noise while retaining speech.

The ideal binary mask has been described as a goal of speech segregation in the field of computational auditory scene analysis (Wang, 2005). Given a priori information of the speech and noise before mixture, a mask value of 1 is assigned to each TF element of the mixture signal where the instantaneous SNR is greater than a threshold value, which is usually 0 dB for the ideal mask. A mask value of 1 indicates that speech is dominant and that the TF unit should be retained in the processed signal. Conversely, a mask value of 0 is assigned if noise dominates the TF unit and, therefore, indicates a TF unit that should be removed from the noise reduced signal.

Generation of the ideal binary mask for a given speech and noise mixture requires a priori access to the separated speech and noise signals. If available, applying the ideal binary mask to the noisy mixture completely recovers the speech signal, demonstrated in both normal hearing (Brungart et al., 2006; Li and Loizou, 2008) and CI listeners (Hu et al., 2007; Hu and Loizou, 2008).
Since the ideal binary mask is used as a target for speech segregation, comparison between an estimated binary mask and the ideal binary mask has been suggested as a method for quantifying performance of noise reduction algorithms. The comparison is described in terms of hit rate (HIT) and false alarm rate (FA). A HIT is assigned for TF elements where both masks are equal to 1 and a FA is assigned when the test mask has a binary value of 1 but the ideal mask has a value of 0. HIT and FA can also be described in terms of type I and type II binary mask error rates, where HIT = 1-type II and FA = type I. Type I errors are interpreted as noise-related errors where type I errors indicate noise in the signal, while type II errors are interpreted as speech distortion-related errors where type II errors indicate speech removal.

Kim et al. (2009) proposed HIT-FA as a metric, and suggested that the relative contribution of HIT and FA might be unequal and could generally be related to the speech material. Kim et al. (2009) suggested that more emphasis be placed on the FA rate, consistent with the speech intelligibility data presented by Li and Loizou (2008), although the errors introduced in that binary mask study were introduced randomly and not simultaneously as type I and type II errors. Consequently, these errors may not reflect the pattern of errors introduced by real noise reduction algorithms. Mauger et al. (2012b) has also followed this approach and suggested that the relative proportion of type I and type II errors reported to be optimal for normal or acoustic hearing (Brungart et al., 2006; Li and Loizou, 2008) is different from the optimal proportion for CI recipients. Mauger et al. (2012b) showed that CI recipients both prefer and obtain improved speech intelligibility when the proportion is weighted in the direction of less type I errors and more type II errors. That is, CI recipients perform better when more noise is removed, even with higher levels of speech distortion as a consequence.

1.2.2 Perceptual measures

Speech intelligibility in the presence of background noise is a primary means for assessing the performance of noise reduction algorithms in human listeners. An improvement in speech intelligibility by introduction of a noise reduction algorithm is a good indicator that an algorithm is likely to be beneficial to the CI recipient. However, other perceptual indicators can be important in describing the benefits of noise reduction. Sound quality, acceptable noise level and ease of listening are common examples. These perceptual responses are usually acquired in the controlled environment of a sound treated room, where the target speech and noise level can be carefully controlled.

Speech reception threshold (SRT)

One approach to assessing speech intelligibility is to measure the recognition of sentences in the presence of background noise at a fixed SNR. The proportion of correct responses can be used to report a percent correct score. However, when scores approach 0% or 100% for any reason, floor/ceiling effects are found, which makes it difficult to detect differences between different conditions. Therefore, the test SNR must be chosen carefully to suit the test material, the test subject, and the sound processing in order to avoid floor and ceiling effects.
An alternative method is to measure speech reception threshold (SRT), which does not suffer from the same floor and ceiling effects as fixed SNR testing. SRT is usually defined as the SNR at which the subject can understand 50% of the material (Levitt, 1978). An adaptive test is used to find the SRT by adjusting the SNR up or down depending on the subject’s response to each item in the test. The adaptive test is designed so that the 50% point can be inferred from the data collected during the test. The main advantage of an adaptive SRT test is the 50% point does not need to be known prior to testing.

A disadvantage of an adaptive SRT test is that the input SNR at which the algorithm is evaluated is controlled by the subject’s response and, by definition, alters the SNR to a point where the subject understands only 50%. However, the SNR at that point may vary with subject and/or algorithm, meaning the subjects and algorithms under evaluation are tested at different input SNRs. This can lead to difficulties in summarising data across subjects or comparing algorithms, especially if the algorithms under evaluation are known to vary in performance as the input SNR changes. An adaptive SRT test cannot be used to evaluate at a specific, predetermined SNR and, therefore, cannot be used to model environments with a specific SNR. Nevertheless, SRT testing provides a fast and efficient means for assessing speech intelligibility in CI subjects, and other tests involving sound quality or acceptable noise level can be used to judge performance at SNRs that are not covered with SRT testing.

**Sound quality (MUSHRA)**

Aside from speech intelligibility, sound quality is an important aspect of noise reduction processing. A technique called MUltiple Stimulus with Hidden Reference and Anchor (MUSHRA) was developed to assess sound quality ratings of degraded speech in telecommunication networks (ITU-R.Recommendation.BS.1534-1, 2003). The task involves making sound quality ratings between different processing conditions by comparing acoustic samples against one another and against a reference and anchor condition. One of the reference conditions is labelled (Reference), and the user is blinded from the processing corresponding to the other conditions, including the hidden reference and anchor, usually by labelling them with letters (e.g. A through F). The user is required to rate the sound quality of each conditions on a scale labelled ‘very poor’, ‘poor’, ‘fair’, ‘good’, and ‘excellent’. The ratings are made on a continuous scale and converted to a score from 0 to 100. A computer is usually used to administer the test, and allows the user to switch quickly between different conditions. The resulting data is amenable to analysis across a group using paired t-tests or repeated-measures analysis of variance.

An important aspect of the task is that the sound quality rating of any particular condition is related to the sound quality of the other conditions to which it is compared, in particular the reference and anchor conditions. Chosen carefully, the reference and anchor provide important end points that enhance the reliability and repeatability of the test. Aside from the reference and anchor, the relative quality of other conditions can influence the absolute sound quality ratings obtained with the MUSHRA test. Zielinski et al. (2007) found that sound quality ratings depend on the range and spread of sound quality samples that are presented during the task, and suggest that analysis of absolute ratings across different configurations of the
MUSHRA task should be done with care to avoid bias effects. As opposed to taking the absolute ratings, Zielinski et al. (2007) suggest ranking the quality ratings, rather than analysing the absolute scores, as a bias free method.

The MUSHRA test was designed for use in telecommunications systems, where processing produces a degradation of sound quality due to coding and decoding the signal. The intention is usually to compare various degraded versions of the signal. The reference signal is the unprocessed original signal, whilst the anchor is a low pass filtered version of the original, representing an extreme case of processing with poorest sound quality compared to the others under evaluation. In CI users, an adapted version of the MUSHRA test was used to evaluate sound quality of music under different high pass filter conditions (Roy et al., 2012). In that study, the authors advocated the MUSHRA test as a user-friendly and versatile tool to measure the effects of a wide range of acoustic manipulations on sound quality perception in CI users.

**Acceptable noise level (ANL)**

The acceptable noise level (ANL), originally called the tolerated signal-to-noise ratio (Nabelek et al., 1991), was introduced to study the toleration of background noise in hearing aid users. A relationship between acceptable noise level and uptake of hearing aids was found which showed that those subjects who tolerated higher levels of noise in the ANL task were more likely to be full-time users of hearing aids (Nabelek et al., 1991).

The task involves listening to running speech in the presence of background noise and adjusting the SNR until the noise is at a level that can be tolerated while still being able to follow the running speech. The task is conducted in two stages, where the first stage involves adjusting the level of running speech to a comfortable level, while the background noise is absent. In the second stage, the noise is introduced and the level of the noise is adjusted to find the acceptable noise level, reported as an SNR.

While established as a test to predict hearing aid uptake, the ANL test has also been used as a means for assessing the benefits of noise reduction in hearing aids. As an alternative outcome measure to traditional speech intelligibility tasks, the relationship between speech intelligibility and ANL is of interest. Mueller et al. (2006) evaluated the effects of digital noise reduction in hearing aid wearers by measuring both ANL and speech intelligibility. Speech reception thresholds were reported using the Hearing In Noise Test (HINT, Nilsson et al., 1994). Although there was no intelligibility improvement on the HINT, digital noise reduction provided a 4.2 dB improvement on the ANL task, indicating that the processing made the background noise more acceptable (Mueller et al., 2006). In other words, there was an improved ease of listening in noisy situations. Additionally, there was no correlation between HINT and ANL, suggesting that the HINT and ANL tasks were involved in measuring different aspects of listening in noise. On the other hand, Freyaldenhoven et al. (2005) evaluated directional microphones in hearing users in a speech in noise task with speech at 0 degrees and noise at 180 degrees. They found directional benefit of 3.5 dB and 3.7 dB in SRT and ANL tasks, respectively. They reported that SRT and ANL benefits were not significantly different from each other and that SRT and ANL benefits were significantly correlated (Freyaldenhoven
et al., 2005). Therefore, it is possible that ANL ratings of hearing aid users are associated with aspects of listening that depend on the signal processing and noise environment in which the evaluation takes place.

The lack of relationship between intelligibility and ANL has been demonstrated in CI users, where no correlation between HINT and ANL scores was found (Plyler et al., 2008). The study with nine adult CI users also included two questionnaires: the Abbreviated Profile of Hearing Aid Benefit (APHAB) and a satisfaction questionnaire, both of which compared pre vs. post implantation conditions. Correlation between ANL and overall satisfaction was demonstrated but there was no correlation found between ANL and APHAB outcomes. To explore this further, Donaldson et al. (2009) performed a similar study in CI users looking for relationships between speech intelligibility (BKB-SIN test), ANL and APHAB. The study had 20 CI subjects, and found that BKB-SIN and ANL were not related to each other. However, BKB-SIN and ANL scores together accounted for 75% of the variance of the global aided APHAB score. BKB-SIN accounted for 33% and ANL accounted for 40% of the variance. This suggests that intelligibility and ANL are not related for CI users, but they are both important metrics which determine a substantial component of the benefit provided by the CI device as reported through the APHAB questionnaire.

### 1.2.3 Acoustic test environments

The performance of noise reduction algorithms is highly dependent on the acoustic situation used for evaluation. In order to study single channel noise reduction alone, the use of spatially co-located target and noise is usually sufficient, even if the real world relevance of such a situation is not high. However, to study multi-microphone algorithms, a spatially separated noise environment is necessary. The specific noise location and/or noise type can be chosen to demonstrate the benefits of a particular algorithm. However, this can lead to tests that have limited real world relevance or that do not provide practical and meaningful results. For example, a fixed directional microphone with a null at 180 degrees tested with target speech at zero degrees and noise at 180 degrees will provide a strong result because the noise is located in the spatial null.

There is a general trend to evaluate noise reduction algorithms in more realistic environments, where the competing noise is modulated, spatially separated and non-static in location. The advantage of pursuing this realism is to create test environments which more closely reflect the outcome users may obtain in the field. Studies that have evaluated multi-microphone systems have used a single spatially separated noise at a fixed location (van Hoesel and Clark, 1995; Hamacher et al., 1997; Wouters and Vanden Berghe, 2001; Stickney et al., 2004; Li and Akagi, 2006; Spriet et al., 2007; Kokkinakis and Loizou, 2008; Li et al., 2008; Chung and Zeng, 2009; Van den Bogaert et al., 2009) or multiple noise sources, spatially separated and at fixed locations (Chung et al., 2006; Li and Akagi, 2006; Spriet et al., 2007; Li et al., 2008; Van den Bogaert et al., 2009). Typically, testing in these environments demonstrates very large performance benefits of multi-microphone algorithms, particularly emphasised when only a single masker location is used, such as the setup depicted in Figure 1-1.
Most multi-microphone algorithms perform extremely well in these idealised conditions, and thus the tests have limited use when trying to discern between different algorithms. This is because adaptive aspects of algorithms cannot be evaluated. Li et al. (2008) used a more complex environment where highly non-stationary, multiple noise sources were used. Three interfering talkers were used, which were spatially separated at fixed locations. Loizou et al. (2009) created a similar setup for evaluating the masking release of bilateral CI users by using up to three independent talker maskers, spatially separated at fixed locations. Chung and Zeng (2009) evaluated directional microphone performance in CI users using a noise environment where the noise locations changed during the test. The noise sources used were steady-state noise with differing frequency shapes. Ricketts and Henry (2002) evaluated directional microphones in hearing aids using spatially separated cafeteria noise with static locations, and also included a moving noise source condition.

Nilsson et al. (2005) describe a spatially separated HINT test using both steady-state and modulated multi-talker noise, although the masker positions are fixed and do not change during the test, as depicted in Figure 1-2. Another commercially available alternative is the R-SPACE restaurant noise (Compton-Conley et al., 2004), which uses a small circle of eight loudspeakers for sound reproduction. The R-SPACE system may not accurately reproduce the intended sound field at the microphone input ports. This is due to the small distance (610mm) between the reproduction loudspeakers and test point, which creates unnatural sound propagation across the test microphones. This is particularly apparent when the test device uses binaurally located microphones and the loudspeakers are in the near field of sound propagation. Due to this restriction, R-SPACE was not considered a suitable long-term solution for our laboratory, and consequently was not used in this study.

Another alternative is to directly record the noisy environment using microphones located in situ, and then replay the noise recordings for later evaluation (Hamacher et al., 1997; Li and Akagi, 2006; 2008). In these situations, recordings are specific to the microphone hardware.

Figure 1-1: Typical speech at 0 degrees with co-located 4 talker babble noise source at 90 degrees (S0N90) e.g. (Spriet et al., 2007)

Figure 1-2: Spatialized HINT test (Nilsson et al., 2005), each noise source is co-located 4 talker babble, 4x4=16 competing talkers in total
used, including microphone location and spacing effects. The recorded signals can later provide input to a speech processing platform but commercial BTE sound processors generally do not have two external input channels to allow this and their use is limited to non-commercial research platforms. Direct recordings of a specific environment are beneficial in that they are true to life; the disadvantage is that the recordings are not easily shared or used as input to a generic speech processing platform. There is the limitation that only specific environments are recorded and tested, and unique qualities of those environments cannot necessarily be generalized more widely.

1.2.4 Field-based assessment

Due to limitations in re-creating true-to-life acoustic environments within the laboratory, field-based assessments are commonly used to evaluate algorithms more widely. Commonly, exposure to the experimental algorithm is accompanied by a questionnaire to assess the effectiveness of the algorithm across a range of subjective metrics. Well-known examples used in the hearing aid industry are the Abbreviated Profile of Hearing Aid Benefit (APHAB, Cox and Alexander, 1995) and the Speech, Spatial and Qualities of Hearing Scale (SSQ, Gatehouse and Noble, 2004). These are questionnaires that are administered after a period of exposure, and necessarily require the user to recall their listening preference and/or experience in different situations when answering the questionnaire. The task requires that the respondent must have experienced the situation, tried the different programs/settings within the situation and come to a decision about preference. In order to report this via the questionnaire, the details need to be accurately recalled at some time after the event, perhaps several days or even weeks after. Consequently, there is inherent variability in responses under these circumstances and large samples sizes may be necessary in order to draw robust conclusions from the data.

1.3 Noise reduction for CIs

The performance of CI users quickly degrades in noisy conditions (Muller-Deile et al., 1995; Dorman et al., 1998; Wolfe et al., 2009). SRTs are typically 10 to 25 dB poorer than their normal hearing counterparts (Hochberg et al., 1992; Wouters and Vanden Berghe, 2001). Therefore, noise reduction plays a crucial role in improving the performance of CI users where both single microphone and multi-microphone systems can be used.

1.3.1 Single microphone noise reduction

Principles of traditional single microphone noise reduction use signal statistics to separate the speech and noise. There is an underlying fundamental assumption that the energy distribution properties of noise and speech are different. Modulation-based algorithms assume speech is more highly modulated than noise, while statistical approaches presume the energy distribution of the speech and noise are Gaussian, or follow other analytic statistical distributions, for example super-Gaussian. These algorithms typically fail to provide intelligibility improvement in acoustic applications for normal hearing listeners. This also
applies to hearing impaired listeners wearing hearing aids, except in very specific noise environments, for example car noise (Hu and Loizou, 2007; Loizou, 2007).

In the application to CI sound processing, the success of single channel noise reduction systems has been quite different to hearing aid applications. In particular, speech intelligibility improvement has been demonstrated with algorithms for CI users that have failed to provide benefit for hearing aid users (Buechner et al., 2010; Hu and Loizou, 2010; Dawson et al., 2011; Hersbach et al., 2012). The maximum benefits for CI users are demonstrated in stationary noise with benefits of around 2.5 dB in SRT and 24 percentage point improvement in word recognition. The benefit of single-channel noise reduction is reduced when the interfering noise is non-stationary, such as when the interfering noise is due to competing talkers (Hersbach et al., 2012).

The contrast with acoustic applications highlights that noise reduction algorithms are able to enhance performance if specifically tailored for CI use. One key difference between CI and hearing aid applications is the relative trade-off between speech distortion and noise reduction that is suitable for CI users. CI users have been found to hear better with more noise reduction (and hence more speech distortion) than is tolerated by hearing aid listeners (Mauger et al., 2012b). In addition, algorithm tuning is usually different for CI applications (Mauger et al., 2012a). A potential reason for the tolerance of CI users could be due to the signal-distorting process of maxima channel selection used as a baseline, such that the distortions introduced by noise reduction do not have a large impact on the already distorted signal.

1.3.2 The binary mask

Single channel noise reduction based on binary masking has attracted attention in the application to CI speech processing (Hu and Loizou, 2008). The objective is to remove components of the signal that have poor SNR whilst retaining those with good SNR. A binary mask is used to remove each TF element with a negative SNR, whilst those with positive SNR are passed on for further processing and ultimately stimulation in the CI. For noise reduction processing, the technique lends itself well to computational training since the ideal mask can be reduced to a binary decision for each TF unit.

Recently, binary classifiers have been trained for the purpose of noise reduction, with the ultimate goal of estimating the ideal binary mask (Kim et al., 2009; Healy et al., 2013). Both of these studies demonstrated speech intelligibility improvements for normal hearing listeners, an achievement which had previously eluded researchers working on single channel noise reduction. Kim et al. (2009) trained a Gaussian Mixture Model (GMM) binary classifier and demonstrated improvement in intelligibility for normal hearing listeners. This was later evaluated in CI users with similar success (Hu and Loizou, 2010). Healy et al. (2013) trained a Deep Neural Network (DNN) binary classifier and demonstrated even better performance, judged by the binary mask estimation errors. Nevertheless, in these studies, the speech material used to train the algorithms was largely the same as the material used for testing. In particular, the specific target talker, the noise type and SNR were common to training and testing phases. Consequently, a practical implementation for noise reduction in a real-time
hearing aid or CI, where prior knowledge of these properties is not known, must estimate these parameters from the input signal. At present, the effects of estimation errors in these properties are not well known, and the area remains an active research topic.

Fink et al. (2012) used a real-time binary mask estimator, and although no improvement was demonstrated for the normal hearing group that were tested, it yielded significant improvement for some of the hearing aid subjects. In contrast, the most impressive benefit was obtained in the CI group who demonstrated a significant improvement of 20% in word recognition rate.

Although other single channel noise reduction techniques do not explicitly estimate the ideal binary mask, it can be useful to analyse them from this view point. This idea is explored further in this thesis by performing theoretical analysis of noise reduction processing from the perspective of binary mask errors. Indeed, any algorithm that provides a noise reduced output equivalent to the ideal binary mask will likely yield positive outcomes. This concept provided significant inspiration for the design of the noise reduction algorithm developed in this research project.

1.3.3 Multi-microphone noise reduction

Many hearing aids and CI sound processors utilize more than one microphone, significantly broadening the scope for noise reduction processing. Sound impinging on the microphone array is subject to a time delay between the microphone inlet ports which is dependent on the direction of arrival. This results in the ability to use information about spatial location of sound sources not possible with a single microphone. A common approach is to use two omni-directional microphones to form fixed or adaptive directional microphones. This is called beamforming, and a common practical implementation is the generalised side-lobe canceller (GSC, Griffiths and Jim, 1982).

The benefit of multi-microphone noise reduction becomes apparent when the target and noise are spatially separated. The physical separation of the microphones provides an acoustic path difference for impinging sound. This difference is dependent on the direction of arrival relative to the microphone array (Figure 1-3). This feature is exploited to spatially filter the signal and reduce the noise. Using small dual-microphone arrays on a single sound processor, large improvements due to adaptive beamforming have been demonstrated for CI recipients. In specific acoustic scenarios involving a single interferer at 90 degrees, improvements of 10 dB SRT on average, were shown for an adaptive beamformer implemented on a body worn sound processor (Wouters and Vanden Berghe, 2001). This algorithm later became known as Beam and 7-16 dB improvement over a fixed directional microphone was shown using an implementation on the Freedom BTE device (Spriet et al., 2007). Chung and Zeng (2009) evaluated an adaptive directional microphone (ADM) which was able to gradually adjust the null position from 90 to 270 degrees demonstrating a 56 percentage point improvement in word recognition over an omni-microphone condition.
Introduction

Recently, three directional settings available in the Cochlear Ltd CP810 processor were evaluated in spatially separated noise, where four independent competing talkers were presented from locations in the rear hemi-field (Hersbach et al., 2012). The location of each competing talker was randomized among seven loudspeakers arranged at 30° intervals from 90° to 270°, and at each three second interval during the test, one of the four competing talkers was randomly chosen and assigned a new random location. In this challenging environment, the benefit demonstrated was 3.7 dB SRT for a fixed directionality pattern called Zoom. A 5.3 dB SRT benefit was demonstrated for an adaptive beamformer called Beam. Both methods were compared to the Standard directionality setting. The superior performance of Beam was due to the adaptive noise cancellation stage, and represents the typical outcome when comparing fixed directional processing to adaptive beamformers; i.e., addition of the adaptive filtering stage resulted in improved performance.

Bilateral systems that utilise microphones located at both ears have also demonstrated benefit for CI users (van Hoesel and Clark, 1995; Hamacher et al., 1997; Kokkinakis and Loizou, 2008; Kokkinakis et al., 2012). However, these bilateral algorithms require a full bandwidth audio link between the two microphones, restricting their implementation in presently available commercial CI systems.

1.3.4 Post-filter algorithms

Single channel algorithms perform best in stationary background noise, whilst multi-channel directional algorithms work best in conditions where the noise location is spatially separated.
from the speech. This implies that there is potential for single channel and multi-channel algorithms to work collaboratively; a single channel approach to suppress diffuse or co-located noise and a directional approach to suppress spatially separated noise. A diffuse noise field is one where sound arrives from all directions, making the signals at the microphones effectively random in terms of determining the direction of arrival of sound. On the other hand, a coherent source is one that has a defined location, resulting in co-ordinated signals at the microphones. In practice, sound is often reflected from multiple objects creating reverberation, resulting in a sound signal that is neither perfectly coherent, not perfectly diffuse.

Hamacher et al. (1997) evaluated a system on CI users using two bilaterally located microphones to drive a single monaural output. A time-domain adaptive beamformer algorithm used for reduction of coherent noise was compared against a frequency-domain adaptive beamformer extended with spectral subtraction for reduction of coherent and diffuse noise. Whilst both algorithms provided a benefit in all conditions tested when compared to the unprocessed condition, the adaptive beamformer extended with spectral subtraction demonstrated a superior improvement in real life cafeteria noise which contained both diffuse and reverberant noise.

In order to enhance beamformer performance, approaches to the design of a so-called post-filter have been explored from a theoretical perspective. Zelinski (1988), Meyer and Simmer (1997), McCowan and Bourlard (2003), and Leukimmiatis et al. (2006) derived optimal filters that differed in the theoretical assumptions made on the sound field. Wolff and Buck (2010) provide a generalised view on microphone array post-filters and provide a description of the theoretical relationships between various post-filter designs. For example, Zelinski (1988) assumed a perfectly diffuse noise field, while McCowan and Bourlard (2003) assumed knowledge of the noise coherence across the microphone array. However, these assumptions generally do not hold, nor are they transferrable to different situations without prior knowledge of the acoustic environment. As a consequence, these theoretical approaches provide limited practical solutions to improve beamformer performance in CI devices.

As an alternative to relying on assumptions of the speech and noise distributions or coherence, Cao et al. (1996) used a post-filter based purely on spatial filtering. The principle of operation was based on a dual-beamformer stage that had a main beamformer aimed at the desired speech source and a reference beamformer designed to pick up all signals except the desired speech. The dual-beamformer outputs were used as speech and noise estimates in a modified Wiener post-filter. In a system with seven omni-directional microphones, they reported on a theoretical measure of segmental SNR, showing an improvement of 9.3 dB over the main beamformer stage. Wolff and Buck (2008) proposed a similar spatial post-filter based on using a spatial noise reference (blocking matrix) as the noise estimate, and described an a posteriori SNR estimate used in a statistically optimized maximum a posteriori (MAP) estimator. In a four-microphone system, they compared the MAP post-filter with a traditional Wiener post-filter. They found substantial improvements in computer word recognition with error rate benefit of between 20 and 40 percentage points depending on the noise type and input SNR used (0-18 dB SNR). A simplified method to that of Wolff and Buck (2008) was proposed by Hegner et al. (2009) which also used the speech and noise reference from the beamformer stage in a
subsequent Wiener filter. However, the filter weights were applied to the output of the adaptive beamformer stage as opposed to the fixed beamformer stage. FFT and wavelet based structures were presented which showed that the simplified method used within the FFT structure failed to produce improvement in a spatially surrounding cocktail party noise according to the segmental SNR that was calculated. In contrast, the wavelet based structure demonstrated some improvement in that noise type. Although evaluated theoretically without involving human listeners, these studies provide evidence that the use of a spatial post-filter can improve beamformer output, and improve it in such a way that is superior to the theoretically optimal Wiener post-filter.

Yousefian and Loizou (2013) proposed a dual-microphone coherence-based spatial filter for CI and demonstrated 5-10 dB SRT improvement over a fixed directional microphone in an anechoic room with one or two competing talkers. The algorithm used the coherence function between the two microphones to show the degree of correlation between two signals at a particular frequency and assumed that speech and noise were coherent across the microphone array. The principle used to separate speech and noise was based on the spatial coherence functions for differential microphones (Elko, 2001). The performance of coherence-based approaches are known to depend to a large extent on the acoustics of the environment such as the room reverberation, the orientation of the microphone array and the spatial distribution of sound sources, as well as the directivity of the microphones used to measure coherence (Martin, 2001b). This dependence may account for the degraded performance in reverberation demonstrated by Yousefian and Loizou (2013) in normal hearing listeners. The benefit was substantially reduced to 0-2 dB SRT when evaluated in a moderately reverberant room (RT60=465 ms) suggesting the coherence assumption was not as strong under this condition.

A dual microphone mask-based design based on phase-errors was proposed by Aarabi and Guangji (2004). The algorithm took into account a pre-determined direction of arrival of the target signal, and calculated the resulting phase difference between the two microphones. The underlying assumption was that large variance in the phase difference was associated with noise and/or reverberant energy, and low variance was associated with the target direction. Based on the phase-error, a mask-based attenuation was applied, improving performance over super-directive beamforming and post-filtering. This approach was later extended to include various masking options and investigated in NH and CI listeners (ur Rehman Qazi et al., 2012). The system utilised bilaterally located microphones and the output was delivered to the test subjects monaurally. The study revealed the phase-error based processing provided significant benefit in competing babble situations compared to the unprocessed signal. The benefit was between 2 dB SRT (S0 N90/180/270) and 8 dB SRT (S0 N90). The different masking options that were evaluated indicated that CI users were more tolerant than NH listeners to distortions in the signal.

A similar spatial filtering approach based on the direct analysis of dual-microphone phase difference was proposed by Goldsworthy et al. (2014) and evaluated in CI listeners. The algorithm used a short term (10 ms) calculation of the phase difference to estimate the direction of arrival (DOA) of sounds and spatially filter the signal, attenuating signal components with DOAs outside a specified target range. The algorithm used microphones
from a single BTE and was evaluated in a room with some reverberation (RT60=350 ms) using separate consonant and vowel recognition in an adaptive SRT procedure. The background noise was time-reversed competing speech from three fixed locations (90, 180 and 270 degrees). The reported benefit of the phase-based spatial filter over an omni-directional microphone was 5.8 to 10.7 dB SRT. The benefit over a fixed directional pattern was 2.2 to 7.0 dB SRT.

1.4 Practical considerations

The controlled environment of the laboratory provides a mechanism for evaluating noise reduction algorithms. However, the sound booth is limited in re-creating true-to-life scenarios that encompass all aspects of a given acoustic environment. Multiple sources of noise, reverberation due to room acoustics and microphone sensitivity matching are important factors known to affect directional microphone performance. Large benefits have been demonstrated using adaptive beamformer algorithms for CI users when there is a single noise source in low reverberation, but that performance benefit is reduced with increased number of noise sources and with increased reverberation (van Hoesel and Clark, 1995; Hamacher et al., 1997; Wouters and Vanden Berghe, 2001). Drift in the sensitivity of microphones within the array leads to a distorted spatial response (Buck, 2002) further impacting performance of array-based noise reduction.

1.4.1 Reverberation

Many real acoustic environments are reverberant, meaning that sound emanating from a particular source arrives at the listener via a direct acoustic path and via paths that reflect off nearby objects and surfaces. Examples of anechoic or low reverberant sound fields include sound treated rooms and outdoor environments where the extent of acoustically reflective surfaces is low. On the other hand, reverberant environments are those where there are many acoustically reflective surfaces that are generally hard and stiff, for example uncovered concrete, brick or tiles. Examples of highly reverberant environments are halls, exhibition spaces, prison cells, etc.

Reverberation time is defined as the time in seconds that it takes for sound in a room to decrease in energy by 60 dB after sudden termination (Beranek, 1988). The metric is usually called the RT60, and is a single figure representing the reverberation time of the entire room. Preferably, it is measured at a variety of source-to-listener locations within the room in order to describe the room more generally. A popular method used to measure the RT60 involves obtaining the impulse response between a source and microphone position and computing the time-reversed integral to obtain the energy decay curve (Schroeder, 1965). The RT60 is then extracted from the energy decay curve by interpolating over the linear region to the point where the energy would have decayed by 60dB from the direct stimulus level.

Another important measure of reverberation is the direct-to-reverberant ratio (Griesinger, 2009). It describes how much energy there is in the direct sound, compared to the sound
which arrives at the listening position after reflecting off surfaces in the room. It is calculated from the impulse response by analysing the energy in the direct part of the response compared to the reverberant tail of the response. The time-point, or split time, used to separate the direct from reverberant energy leads to metrics that have different interpretations regarding clarity, speech intelligibility and localisation abilities in the reverberant space. For example, C80 (split time of 80 ms) is used to predict articulation and clarity of different types of music, and provides insight into the musical performance of the room. C50 (split time of 50 ms) is used to predict speech intelligibility, and C7 (split time of 7 ms) used for localisation (Kuttruff, 2009). Whilst all these measures describe the direct-to-reverberant ratio with different split times, a direct-to-reverberant ratio metric (DRR) is usually used to describe a split time of a few milliseconds, over the range from 2.5 -7 ms. The critical distance is also useful as it describes the distance at which the DRR is 0 dB, when the direct and reverberant energy are equal. Listening distances smaller than the critical distance have a higher direct sound energy, while greater distances contain more reverberant energy.

Room acoustics can dramatically alter the perception of sound, and while early reflections are thought to contribute favourably to speech intelligibility, high levels of reverberation ultimately have a negative effect on understanding. This is particularly important for children whose tolerance of reverberation is not as high as adults (Neuman et al., 2010) and for CI users who are more severely affected by reverberation than their normal hearing counterparts (Neuman et al., 2012). The effects of reverberation on CI performance have been studied with and without background noise present. Neuman et al. (2012) showed that reverberation time of 0.8 s led to reduced performance of CI children in an assessment without background noise, while Kokkinakis et al. (2011) demonstrated an exponential decay in performance in CI adults with reverberation time from 0.3 s to 1 s (without background noise). The addition of background noise with reverberation time of 0.6 s and 0.8 s has been shown to further reduce intelligibility for CI users (Hazrati and Loizou, 2012). Similar findings have also been demonstrated in vocoder-based normal hearing studies (Poissant et al., 2006; Whitmal and Poissant, 2009; Tillery et al., 2012) that generally indicate CI performance is expected to decrease with increased levels of reverberation.

High levels of reverberation have an impact on speech intelligibility for CI users, and in addition, reverberation also affects the ability of noise reduction algorithms to enhance the signal. In particular, adaptive beamformer performance is reduced at high levels of reverberation (Greenberg and Zurek, 2001), which has been demonstrated in the application to CI processing (van Hoesel and Clark, 1995; Hamacher et al., 1997; Kompis et al., 2008). Directional microphone benefit in hearing aid users was studied by Ricketts and Hornsby (2003) in different levels of reverberation and target-to-listener distances. The study found that target-to-listener distance reduced directional benefit in moderate reverberation (RT60 = 0.9 s) but did not alter directional benefit in low reverberation (RT60 = 0.3 s). The data supported the use of directional microphones in moderate reverberation as an effective means of noise reduction. The benefit was due to the attenuation of the direct energy of the nearby noise sources (Ricketts and Hornsby, 2003).

Given the detrimental effects of reverberation, speech enhancement algorithms need to be robust to the effects of reverberation encountered in common listening situations in order to
provide benefit in noisy, reverberant conditions. Therefore, an understanding of algorithm behaviour across a range of reverberant conditions was studied within this project during evaluation with CI users.

### 1.4.2 Microphone mismatch

Directional or multi-microphone signal processing by its very nature exploits the differences between microphones resulting from the physical separation of the microphone ports. Sound waves arriving at different ports of the microphone array are subject to time delay and level differences, dependant on the direction of arrival, physical port spacing and surrounding acoustic surfaces like the head and ear. Therefore, the design of signal processing strategies that exploit these differences must make some assumptions about the properties of the microphones themselves, in particular the matching of phase and amplitude characteristics over the frequency range of interest.

Manufacturers that supply microphones to the hearing instrument market sell microphones in matched pairs for this use, with sensitivity matching accuracy in the range of 1-2 dB at 1 kHz and phase matching accuracy in the range of 0.5-1.5 degrees at 200/250 Hz (Knowles, Sonion). In addition to the matching accuracy provided by the microphone manufacturer, the hearing instrument manufacturer has the opportunity to calibrate the assembled product, further increasing the accuracy of matching. However, during normal use, it is typical for the microphone ports and covers to be subject to dirt, moisture and sweat, which can adversely affect the performance of the microphones, ultimately changing the matching accuracy.

Under conditions where ideally matched microphones are not met, directional performance can be affected. Buck (2002) studied the effect of microphone mismatch on fixed directional microphones by analysing the resulting directivity index and response across frequency. The study showed that low frequency directivity was most severely affected by sensitivity mismatch, and that sensor calibration to improve matching could lead to improved performance. For fixed directional microphones, the reduced directivity index of the spatial pattern is expected to result in reduced noise reduction performance due to the reduced null depth. This has implications for adaptive beamformers, which use a speech reference (typically a forward facing directional pattern) and noise reference (typically a rear facing directional pattern) as the directional patterns for these will be altered. In particular, the reduced null depth of the noise reference allows speech from the target direction to corrupt the noise reference signal, so-called speech leakage. Tashev (2005) reported the effect of microphone magnitude and phase mismatch on beamformer noise gain. Magnitude had a greater effect than phase difference and a 3 dB mismatch increased the noise gain by about 4 dB. In order to provide good performance, an ensuing recommendation was provided, that microphones should be matched to within +/- 0.5 dB (Tashev, 2005).

Various solutions have been proposed to address microphone matching issues. Self-correction algorithms that attempt to adaptively match the microphones have been proposed (Buck, 2002; Buck et al., 2006), which offer a means for mitigating the negative effects of mismatch. Another solution has been proposed that addresses the issue of speech leakage in the noise reference for a multi-channel Wiener filter (Doclo et al., 2007). The algorithm introduced a
parameter that explicitly allowed the system to tune the trade-off between noise reduction and speech distortion. The parameter can be chosen to protect the target from distortion (at the expense of less noise reduction), thus mitigating the effects of speech leakage in the noise reference which can be caused by microphone mismatch.

Whilst the impact of microphone mismatch on directional performance has been predicted from a theoretical viewpoint, there is a lack of clinical data with CI users describing the impact on performance. It is not yet clear how to translate bench-top theoretical measures of directivity index or noise gain into clinical performance, particularly for adaptive algorithms that have spatial patterns which change over time. When does microphone mismatch start to affect CI performance? How much mismatch is too much? When should the microphones be recommended for service/replacement? These questions remain largely unanswered in the context of directional and multi-microphone sound processing for CI users.

1.5 Outline of the thesis

The main research objectives and chapter by chapter overview of the thesis are provided in the following sections.

1.5.1 Main research objectives

The main research objectives of this thesis project are:

1. To develop a noise reduction algorithm novel to CIs that provides better performance in noisy conditions than existing techniques.

2. To evaluate the noise reduction algorithm in a range of acoustic conditions that are representative of situations encountered during typical normal use of a CI. The goal is to obtain an extensive understanding of the algorithm and to determine how changes to algorithm parameters affect performance.

3. To develop a system that allows CI users to provide direct feedback on their listening experience whilst on location in the field. The goal is to obtain integral information about listening preference in the field that is difficult to obtain via a questionnaire.

To achieve these objectives, evaluations with CI users are performed, primarily with speech intelligibility tasks, but also with sound quality and acceptable noise level tasks. Existing commercial algorithms from Cochlear Ltd are used as baseline algorithms in order to compare performance to state of the art in CI processing. Evaluation is performed in a range of competing noise scenarios, involving multiple noise sources, reverberation and microphone mismatch. Ultimately, field-based preference is obtained from CI users able to provide on-the-spot voting using their sound processor’s remote control.
1.5.2 Chapter by chapter overview

In Chapter 2, the noise reduction algorithm is introduced and the principles of operation are discussed. This is documented through block diagrams and mathematical operations as well as an accompanying theoretical analysis of operation. Directional sensitivity plots are provided for the algorithm, as well as those for baseline algorithms, Standard, Zoom and Beam. The influence of three algorithm parameters is also described from a theoretical viewpoint, as it is the variation of these parameters that forms the basis of perceptual evaluation in Chapters 4-8.

Chapter 3 is dedicated to the experimental methods used in the experimental Chapters 4-8 that follow. Both theoretical performance measures and perceptual outcome measures that are used in experiments are described as well as details of the noise environments used for evaluation. The voting paradigm used in Experiment V (Chapter 8) is also described, including methods of data extraction and post processing used in the voting experiment.

An initial performance evaluation with CI listeners is performed in Chapter 4. The main objectives of this experiment are to identify the intelligibility benefit due to the algorithm, if any, and to study the effect of the bias parameter used to adjust the algorithm’s aggressiveness. The algorithms’ bias parameter is varied, and performance is compared against the Beam baseline algorithm through speech intelligibility testing in one type of competing noise (4-TB-rear). Theoretical performance measures are used in order to guide the range of bias parameter values used in the clinical evaluation with CI users.

In Chapter 5, the range of noise types used for evaluation are expanded by introducing more competing talkers (20-TB-full) and performing tests in a room with more reverberation. The bias parameter is again varied as well as the parameter related to signal smoothing. Both speech intelligibility and sound quality are assessed in two types of noise (4-TB-rear, 20-TB-full) and testing in quiet is also performed to highlight any negative, target cancellation-related deficiencies known to exist for adaptive beamformers. The number of baseline algorithms is expanded to include Standard, Zoom and Beam. As with the previous experiment, theoretical performance is evaluated in order to refine the set of parameter values used in the clinical evaluation.

In order to further explore the effects of reverberation, Chapter 6 studies a range of direct-to-reverberant ratios by changing the distance of the target and noise from the listener and, through the use of a room simulator, extending to more extreme reverberation. For this evaluation, the algorithm parameters are not varied but are rather fixed, and an additional baseline condition using an omni-directional microphone is introduced. Speech intelligibility is assessed through SRT testing in one noise type (4-TB-full).

The effects of microphone mismatch are studied in Chapter 7 by introducing a broad band level difference between the two microphones, known to distort the spatial response. Similar to Chapter 6, algorithm parameters are fixed and speech intelligibility is assessed in a single noise type (4-TB-full) although sound quality is also assessed. Baseline conditions are Standard, Zoom and Beam, and theoretical performance is assessed to guide the range of microphone mismatch used in the clinical evaluation.
Chapter 8 provides the details of the final experiment in this project, in which laboratory-based evaluations are accompanied by field-based assessment. For this purpose, the algorithm is implemented on a BTE sound processor so that subjects could evaluate the algorithm during take-home use. In this experiment, the focus changes to an understanding the effect of the maximum attenuation parameter used to smoothly transition the algorithm from on to off. In an attempt to gain full understanding of the algorithm, laboratory-based assessment expands the number of noise types to four (2-SWN-rear, 4-TB-rear, 4-TB-full and 20-TB-full) and expands the outcome measures to include speech intelligibility (SRT), sound quality (MUSHRA) and acceptable noise level (ANL). In addition to varying the maximum attenuation parameter, performance is evaluated in combination with a single-microphone noise reduction algorithm. Acclimatisation to the algorithm over time is evaluated and the BTE implementation is compared against the PC-based implementation used in all previous experiments. During the accompanying field-based assessment, subjects voted for their preferred program using their speech processor’s remote control, allowing the evaluation of different maximum attenuation parameters as well as comparison of baseline conditions (Standard, Zoom, Beam, and SNR-NR) against one another.
Chapter 2

Algorithm Design

In this chapter, the design principles of the noise reduction algorithm are explained through block diagrams and mathematical equations. A theoretical analysis of the algorithm behaviour is provided, as well as the variation in behaviour achieved through the modification of three different algorithm parameters.

2.1 Introduction

The design and development of the noise reduction algorithm (SpatialNR) was inspired by ideal binary mask studies in CI users. The ideal binary mask is derived from prior knowledge of the pre-mixed speech and noise signals and, if the SNR is known, the speech can be completely recovered from the noise. In a real-time noise reduction system, the SNR needs to be estimated for each TF unit and single channel systems have delivered some success in CI users, especially in steady state noise.

As an example, the Nucleus 6 CP900 sound processor from Cochlear Ltd has a single channel noise reduction algorithm called SNR-NR (Dawson et al., 2011; Hersbach et al., 2012). The algorithm estimates the noise floor using a minimum statistics recursive smoothing algorithm (Martin, 2001a) and is followed by an a priori SNR estimation stage. A gain application stage based on a parametric Wiener filter is used to attenuate noisy channels. The algorithm was recently evaluated in combination with various directionality configurations and was shown to provide speech intelligibility improvement in stationary, speech-weighted noise, but not in modulated, multi-talker babble scenarios (Hersbach et al., 2012).

The main driving goal of the current work arose from the question; “How can the SNR estimation be improved?”

From this point, and with the use of spatial filtering properties of multi-microphone processing, the SpatialNR algorithm was designed as a means of estimating the SNR, to either replace or further enhance the SNR estimate derived from a single channel approach. The
underlying rationale for determining speech from noise was spatial separation, where information about it was provided through the multi-microphone array.

The SpatialNR algorithm was formulated as a beamformer post-filter designed to enhance intelligibility of the beamformer output. It was designed for real-time, low complexity implementation for a small separation (< 2 cm) dual-microphone array typically used in hearing aids and CIs. No assumption was made about the underlying speech and noise power distributions, and no a priori knowledge of the speech and noise signals was used in the processing. However, the target spatial location was assumed to be in front of the listener.

The algorithm estimated the SNR by analysing the signals from front-facing and rear-facing fixed directional microphones. SNR was estimated from the front-to-back ratio in each frequency band. When more energy was in the front-facing signal, the SNR was presumed to be positive, and when more energy was in the rear-facing signal, the SNR was presumed to be negative. The SNR estimate was used to control the attenuation of noisy channels via a parametric Wiener-like gain function with adjustable gain. The final design was a spatial post-filter that attenuated sound from behind and beside the listener while passing sounds from the front. The final design was similar to the spatial post-filter proposed by Cao et al. (1996) which also used a main beamformer and reference beamformer as inputs to a modified Wiener filter. The main differences were the design of the beamformer stages due to the number of available microphones, the signal smoothing stage, and parameterisation of the modified Wiener filter. Cao et al. (1996) did not provide details on the proper choice of parameter values. However, they speculated that the choice might be related to the acoustic environment, noting that the choice could impact the naturalness of speech in the processed signal and that experimentation for the choice of parameters is of importance.

The Nucleus 5 CP810 and Nucleus 6 CP900 are the two most recent sound processors available from Cochlear Ltd and were used as baseline processing throughout this thesis project (Mauger et al., 2014). The systems have two omni-directional microphones that are used to form three different directional patterns, configurable through the commercial fitting software used to program the device. The first is Standard, offering a fixed directional microphone with moderate directionality. The second provides a fixed Zoom pattern offering high directionality, and the third is an adaptive beamformer called Beam providing an adaptive highly directional microphone response (Spriet et al., 2007). Spatial sensitivity plots of these directional configurations are shown later in Figure 2-7.

### 2.2 Algorithm kernel

The CP810/CP900 signal path and directionality infrastructure used for integration of the SpatialNR algorithm is shown in Figure 2-1 while the algorithm kernel itself is shown in Figure 2-2.
2.2 Algorithm kernel

Figure 2-1: Signal path integration used to evaluate the SpatialNR algorithm. The speech reference (Standard or Zoom) and noise reference signals are used as inputs to Beam and SpatialNR algorithms. A switch provides selection of either Beam or SpatialNR.

Two omni-directional microphone signals (mic1 and mic2) are used as inputs to a spatial pre-filter stage, where two spatial patterns are formed – a front-facing speech reference and a rear-facing noise reference. The CP810/CP900 spatial pre-filter provides the option to choose the speech reference configuration to be Standard or Zoom. The Zoom speech reference is configured to have a front-facing polar pattern of maximum directivity index when worn on the head and therefore provides a better speech reference for the SpatialNR algorithm. The noise reference is configured to have a rear-facing polar pattern with a null in the target, or “look” direction, in this case directly in front of the listener. The spatial sensitivity of the speech and noise references are shown in Figure 2-3 for microphones worn on the left ear of a Knowles Electronics Manikin for Acoustic Research (KEMAR, Burkhard and Sachs, 1975). Note there is increased sensitivity at around 45° on the same side of the manikin head where the microphones are worn, due to the head shadow. The spatial pre-filter signals are created with a delay-and-sum structure, by filtering and summing the microphone input signals in the time domain. The same speech and noise reference signals are also used by the adaptive Beam algorithm, which uses the Standard configuration as the speech reference in CP810 and CP900 devices.

An FFT is used to transform the signals into the complex spectral domain, creating frequency domain representations of the speech reference, $S_k[n]$, and noise reference, $N_k[n]$, where $k$ is the frequency index and $n$ is the time index of overlapping FFT windows. The speech reference and noise reference signals are transformed to the dB domain and filtered separately using first-order IIR filters with independent attack, $0 \leq \beta_A \leq 1$, and release times, $0 \leq \beta_R \leq 1$. 
Figure 2-2: SpatialNR algorithm kernel. The two microphone signals were used to create two spatially filtered signals; a forward-facing speech reference \(S^{dB}\) and a rear-facing noise reference \(N^{dB}\). The SNR \(\xi^{dB}\) was calculated from the ratio of the smoothed speech and noise reference signals, which was used to calculate the noise reduction gains using a parametric Wiener-like function.

\[
\begin{align*}
\overline{S^{dB}}_k[n] &= \beta_S S^{dB}_k[n] + (1 - \beta_S)\overline{S^{dB}}_k[n - 1], \\
\beta_S &= \begin{cases} 
\beta_A, & \overline{S^{dB}}_k[n] > \overline{S^{dB}}_k[n - 1] \\
\beta_R, & \text{otherwise}
\end{cases} \quad (2-1)
\end{align*}
\]

\[
\overline{N^{dB}}_k[n] = \beta_N N^{dB}_k[n] + (1 - \beta_N)\overline{N^{dB}}_k[n - 1],
\]

\[
\beta_N = \begin{cases} 
\beta_A, & \overline{N^{dB}}_k[n] > \overline{N^{dB}}_k[n - 1] \\
\beta_R, & \text{otherwise}
\end{cases} \quad (2-2)
\]

Smoothing in the dB domain is used because it relates more closely to perceptual loudness. The smoothed signals are used to estimate the instantaneous SNR, \(\xi_k[n]\), at each time point, \(n\), and in each frequency band, \(k\),

\[
\xi^{dB}_k[n] = \overline{S^{dB}}_k[n] - \overline{N^{dB}}_k[n]. \quad (2-3)
\]

The SNR estimate \(\xi_k[n]\) is then used as the primary means to attenuate TF bins that are considered noisy. It is used to control a parametric Wiener-like gain, \(H_k[n]\), with adjustable bias, \(\alpha > 0\),

\[
H_k[n] = \frac{\xi_k[n]}{\alpha + \xi_k[n]}, \quad (2-4)
\]

\(H_k[n]\) is a smooth function of SNR and was chosen in preference to a binary gain function due to superior sound quality when used in single-microphone noise reduction for CI recipients (Dawson et al., 2011). The clean signal is estimated, \(\hat{X}_k[n]\), by applying the filter gain, \(H_k[n]\), to the output of the beamformer stage,

\[
\hat{X}_k[n] = H_k[n]S_k[n]. \quad (2-5)
\]
Figure 2-3: (A) Single source spatial sensitivity of Speech Reference (Zoom directionality) and Noise Reference signals and (B) the resulting single source SNR $\xi_{dB}$ according to Equation 2-3. Sensitivity was recorded with a broadband speech-shaped noise stimulus presented under anechoic conditions with device worn on the left ear of a KEMAR manikin. The relative sensitivity is the RMS value across the duration of the stimulus averaged across all frequency channels.

Finally, the output signal, $Y_k[n]$, is formed from a weighted combination of the speech reference signal, $S_k[n]$, and the estimated clean signal, $\hat{X}_k[n]$, using the maximum attenuation parameter, $\gamma$, to mix the two together. This parameter allows the output signal to completely disable ($\gamma = 0$) or completely enable ($\gamma = 1$) the noise reduction processing, with a continuous and smooth transition between the two.

$$Y_k[n] = \gamma \hat{X}_k[n] + (1 - \gamma)S_k[n].$$  \hspace{1cm} (2-6)
2.3 Theoretical analysis

The spatial response or directivity pattern of the SpatialNR algorithm depends on the input signal and changes over time. The spatial location and level of the noise and the target signal, the spatial pre-filtering stage, and the bias parameter, $\alpha$, all have an influence on the system’s spatial response.

By way of illustration, it is useful to consider a single sound source impinging on the array. In this case, the value of $\xi$ is determined by the spatial location of the single sound source in relation to the spatial responses in the pre-filtering stage, specifically the sensitivity of the speech reference compared to the noise reference at that location. Figure 2-3 shows the relative sensitivity of the speech reference and noise reference and the associated $\xi^{dB}$ that is calculated according to Equation 2-3.

In contrast to the single source case, when two signals are present at two different locations, each will contribute to $\xi^{dB}$ by an amount dependent on the level and location of each, and the final $\xi^{dB}$ is calculated from the SNR contributions due to the sound at each location. This is illustrated in Figure 2-4, which shows $\xi^{dB}$ as a function of input SNR for a target signal at 0° and a single noise from different angles ranging from 30° to 180°. For example, consider the case where the noise source is at 90°. In that case, the target source at 0° will contribute a single source spatial SNR of $\xi^{dB} = +32$ dB, and the noise source at 90° will contribute a single source spatial SNR of $\xi^{dB} = -3$ dB.
2.4 Algorithm parameters

There are three parameters that can be adjusted to change the behaviour of the SpatialINR algorithm. The smoothing parameter, $\beta$, changes the smoothing filter time constants (Equations 2-1 and 2-2). The bias value, $\alpha$, of the parametric Wiener filter changes the aggressiveness of noise reduction (Equation 2-4). The maximum attenuation parameter, $\gamma$, controls the post mixing of the estimated clean signal and the noisy input signal (Equation 2-6).

2.4.1 Signal smoothing

The signal smoothing parameter, $\beta$, is used to control the first order IIR smoothing filters with separate attack and release times. The effect of the parameter, which controls the exponential smoothing, is shown in Figure 2-5. A time constant of 10 ms for example has a value of $\beta=0.185$ given the 489 Hz rate of successive FFT frames. The speech reference and noise reference are smoothed independently and with the use of asymmetric smoothing (fast attack and slower release) rapid onsets in either signal could be closely followed. The perceptual effect of altering the smoothing is related to sound quality and the responsiveness of the algorithm to changes in SNR as the speech and noise signal modulate up and down. This was investigated in Experiment II.
Figure 2-6: Parametric Wiener filter which maps Input SNR, $\xi$, to Gain. The effect of bias parameter, $\alpha$, is shown for a range of values.

2.4.2 Bias parameter

The bias parameter, $\alpha$, is used in the calculation of noise reduction gain, $H$ (Equation 2-4). It is used to determine the SNR threshold at which elements are considered noisy and, therefore, removed. The effect of the bias parameter value on gain is shown in Figure 2-6.

The polar response of the SpatialNR algorithm and the effect of bias parameter are shown in the right panel of Figure 2-7. The polar response was measured using a broadband speech shaped noise stimulus presented from each location under anechoic conditions with the device worn on the left ear of a KEMAR manikin. The system transfer function was measured from input to output, and thus represents the whole system response for each processing condition. The SpatialNR bias parameter has the effect of altering the beam width in the target direction as well as changing the attenuation of signals outside the target area. In this way it provides a mechanism to alter the aggressiveness of noise reduction.

Also shown in Figure 2-7 is the response for Standard, Zoom and Beam processing. The interpretation of these polar plots must take into consideration that they measure the sensitivity in each direction at different instances, when only a single sound source is present. That is, the measurement at each angle is made by presenting a single source at that angle, and each angle is measured separately. The polar plot is formed by combining the measurement at all angles into a single figure. Therefore, they do not represent an instantaneous view of the spatial sensitivity in multiple directions simultaneously. Polar plots can be a useful analytic tool, but can only partly describe the behaviour of a system that is dynamic and do not adequately describe polar patterns that change over time. For this reason, the effective noise reduction capability of Beam and SpatialNR, in particular, cannot be easily inferred from the polar plots since the algorithms change over time and respond differently when multiple sources are present.
2.4 Algorithm parameters

Figure 2-7: Single source spatial sensitivity in anechoic conditions of Standard, Zoom, Beam and SpatialNR processing including effect of SpatialNR bias parameter, $\alpha$.

2.4.3 Maximum attenuation

The purpose of the maximum attenuation parameter, $\gamma$, was to provide a smooth transition from no noise reduction to full noise reduction. When the parameter value was 1, only the estimated clean signal was used. When the value was 0, only the noisy input signal was used. For values in between, the two signals were progressively mixed, allowing a smooth transition.

The maximum attenuation parameter obtained its name from the impact it has on the limited gain function that results using an alternative formulation. Substituting Equation 2-5 into Equation 2-6

$$Y_k[n] = \gamma H_k[n]S_k[n] + (1 - \gamma)S_k[n]$$  \hspace{1cm} (2-7)

$$Y_k[n] = [\gamma H_k[n] + 1 - \gamma]S_k[n]$$  \hspace{1cm} (2-8)

The term $\gamma H_k[n] + 1 - \gamma$ represents the gain to be applied to the input signal, and when Equation 2-4 is substituted, it represents a gain function with limited attenuation, plotted in Figure 2-8

$$\frac{\gamma S_k[n]}{\alpha + S_k[n]} + 1 - \gamma$$  \hspace{1cm} (2-9)

In the first four experiments, the maximum attenuation parameter, $\gamma$, was held constant at a value of 1, equivalent to maximum possible noise reduction. In the final experiment (Chapter 8), the perceptual effect of the maximum attenuation parameter was investigated.
Figure 2-8: Gain function ($\alpha = 0$ dB) with attenuation that is limited by the maximum attenuation parameter, $\gamma$.

2.5 Summary

The SpatialNR algorithm has been described in terms of the principles of operation. The system is used to attenuate signals behind and beside the listener while the target position is presumed to be in front of the listener. At its core, the algorithm generates an estimate of the SNR based on the front-to-back ratio of two directional microphones; one forward-facing, another rear-facing. The SNR estimate is used to attenuate noisy spectral channels through the user of a parametric Wiener-like gain function, which provides a mapping of SNR estimate to gain. Three parameters are available to control operation of the algorithm, two of which modify the gain function, while the third is used to control signal smoothing.
Chapter 3

Methods

This chapter describes the experimental methods employed over a series of five experiments that were designed to evaluate the performance of noise reduction algorithms. Perceptual listening tests with CI users were the primary means for evaluation. Theoretical performance measures were used to guide early algorithm development and to help refine the number of test conditions in perceptual testing.

3.1 Introduction

The experiments conducted were centred upon the evaluation of the SpatialINR algorithm as a means for noise reduction. To achieve this, theoretical measurements and listening tests with CI users were conducted. In total, five different experiments were conducted, which explored different parameter settings, different implementation platforms and various acoustic environments.

While the main focus was on clinical evaluation with CI users, theoretical evaluations are presented as part of some experiments, and were used in the early stages of algorithm development to help guide the range of algorithm parameters put forward for clinical testing.

To evaluate noise reduction performance, a controlled sound environment was used that contained target speech and competing noise. In particular, the competing noise source and spatial locations of the competing noise were controlled to create a number of different test environments. Additionally, the sound environment was created in different rooms that had different levels of reverberation.

Sound processing of the SpatialINR algorithm was executed on a PC-based real-time processing system for Experiments I-IV (Chapters 4-7), and finally in Experiment V (Chapter 8), a wearable BTE processor was used. The five different experiments were each conducted with a group of CI subjects. The experimental protocols were designed for statistical analyses of group data in order to draw conclusions for the adult CI population. The following section describes methods
and materials used in the experiments, although not all experiments made use of all methods and/or materials.

3.2 Theoretical performance measures

Two theoretical performance measures were calculated as ways to screen algorithm parameters and refine the number of processing conditions used in clinical trials with CI users. The two metrics that were used were ideal binary mask errors and normalised correlation speech-based STI. The methods for calculation of the metrics are described below.

Each experiment was subject to error rate/STI testing to give an indication of how theoretical performance varied with parameter changes. Error rates were used as a guide to understand the impact of parameter variation on the amount of noise reduction (type I errors) and speech distortion (type II errors). For most parameters, it was expected that a variation in value was likely to lead to an increase in one error and a decrease in the other. This highlights the inherent trade-off involved in tuning the parameter space.

Error rate testing was also used to validate the expected operation of each algorithm on the real-time system. This was done to ensure that the intended changes and switching resulted in the expected change in processing as indicated by the change in error rates. In this way, different switching methods (fast/slow) were verified, both on the real-time system and the CP900 implementation.

Error rates and STI metrics were calculated using off-line processing of input signals with BKB-like sentences (Bench et al., 1979) in the presence of different types of background noise, as used for the clinical evaluation. For each background noise type, the stimulus was 30 s duration, and consisted of BKB-like sentences separate by 1 s gaps. The target sentences were embedded in background noise that changed in its spatial configuration throughout the stimulus. The spatial configuration of the noise changed in a random but repeatable fashion, with changes every 3 seconds, as detailed in section 3.4.2. Signals were generated using the impulse response method, and modelled the same acoustic conditions used for clinical evaluation of SRTs.

3.2.1 STI

As an alternative to traditional STI, a normalised correlation speech-based version of the STI was calculated as means of predicting performance (Goldsworthy and Greenberg, 2004). In addition to predicting the effects of linear and non-linear sound processing operations, the metric is expected to accurately predict speech intelligibility for CI users. This is due to the similarities between CI processing and the calculation of the metric (Goldsworthy and Greenberg, 2004). Specifically, the analysis of spectral envelopes and the frequency bands used for analysis can be made identical.
3.2 Theoretical performance measures

3.2.2 Error rate calculation

Ideal binary mask errors were calculated as means for determining noise reduction performance. According to errors in the ideal binary mask, type I and type II errors were calculated. As opposed to combining the errors into a single metric, type I and type II errors were maintained as separate entities. This is because HIT = 1-type II, FA = type I and HIT-FA has previously been proposed as a metric in the context of binary masking studies but HIT and FA might contribute differently to overall performance (Kim et al., 2009).

In order to calculate error rates, the output signal was analysed after processing with various sound processing settings. This occurred most commonly after changes to a particular parameter value. Errors arose from comparison with ‘ideal’ processing, which contained only the target signal and no noise. This criterion of ‘ideal’ differs from that defined by Wang (2005) who used an SNR-based local threshold determined by a target and noise mixture. The alternative definition was used here because it is independent of the competing noise.

To calculate errors, the ‘ideal’ signal, \( X \), was first processed and the stimulation output signal in units of clinical current levels, \( X_{CL}[n,k] \), was categorised into speech present and speech absent segments to create a speech presence matrix, \( p[n,k] \) (Equation 3-1). Segmentation into TF units was performed across time, \( n \), and frequency, \( k \), in accordance with default settings for Cochlear Nucleus device. That is, 22 frequency channels at 900 Hz. The dynamic range of \( X_{CL}[n,k] \) was assigned from 0 to 200 current levels, using an acoustic input dynamic range of 40 dB. Current levels represented the signal magnitude. Mapping from units of dB to CL was approximately linear over the range. A current level of zero was assigned for all TF units below the lower limit of the input dynamic range. Therefore, all TF units with current level greater than zero were assumed to represent the ideal signal.

\[
p[n,k] = \begin{cases} 
1, & X_{CL}[n,k] > 0 \\
0, & \text{otherwise} 
\end{cases} 
\]  

(3-1)

Next, the noisy signal, \( Y \), was processed to form the noise reduced output signal in units of current levels, \( \hat{X}_{CL}[n,k] \). Errors were calculated by comparing the ‘ideal’ and processed signals. As opposed to assigning binary errors of value 0 or 1, partial errors, \( e[n,k] \), were calculated based on the difference in current levels, \( E_{CL}[n,k] \), between the two signals (Equation 3-2). The difference was converted to dB using the known mapping of acoustic to electric dynamic range. That is, 40 dB acoustic dynamic range was mapped to 200 current level electric dynamic range (Equation 3-3). If the difference between the ‘ideal’ and processed signals, \( E_{dB}[n,k] \), was greater than a threshold of 15 dB, a whole error value of 1 was assigned. For differences smaller than the threshold, a partial error was assigned in a manner that linearly mapped the difference in dB to error rates between 0 and 1 (Equation 3-4).

\[
E_{CL}[n,k] = |X_{CL}[n,k] - \hat{X}_{CL}[n,k]| 
\]  

(3-2)

\[
E_{dB}[n,k] = 40 \times \frac{E_{CL}[n,k]}{200} 
\]  

(3-3)
Methods

\[
e[n,k] = \begin{cases} 
0, & E_{db}[n,k] = 0 \\
\frac{E_{db}[n,k]}{15}, & 0 < E_{db}[n,k] < 15 \\
1, & E_{db}[n,k] \geq 15
\end{cases}
\]

Type I errors, \(e_i[n,k]\), were calculated from the speech absent segments (Equation 3-5). By definition, the ‘ideal’ stimulus level was zero for speech absent segments. Therefore, an error of 1 was assigned to all segments that contained a stimulus level greater than 15 dB above T-level. Partial errors were assigned for stimulus levels between 0 and 15 dB according to the linear mapping.

\[
e_i[n,k] = e[n,k](1 - p[n,k])
\]

Type II errors, \(e_{II}[n,k]\), were calculated from speech present segments (Equation 3-6), with a threshold of 15 dB to achieve a whole error value of 1. If the stimulus level of the processed signal was 15 dB greater or less than the ‘ideal’ signal, an error value of 1 was assigned. Partial errors were assigned when the absolute difference was less than 15 dB according to the linear mapping. The direction of difference was treated equally because the absolute difference between ‘ideal’ and processed signal was used (Equation 3-2).

\[
e_{II}[n,k] = e[n,k]p[n,k]
\]

Error rates were then summed across time and frequency, disregarding the first three seconds of stimulus to allow for initial conditions to settle and algorithm adaptation to occur. Type I error rate, \(E_i\), was reported as the sum of type I errors divided by the number of speech absent segments (Equation 3-7). Type II error rate, \(E_{II}\), was reported as the sum of type II errors divided by the number speech present segments (Equation 3-8).

\[
E_i = \frac{\sum_{n=1}^{N} \sum_{k=1}^{K} e_i[n,k]}{\sum_{n=1}^{N} \sum_{k=1}^{K} (1 - p[n,k])}
\]

\[
E_{II} = \frac{\sum_{n=1}^{N} \sum_{k=1}^{K} e_{II}[n,k]}{\sum_{n=1}^{N} \sum_{k=1}^{K} p[n,k]}
\]

3.3 Laboratory-based perceptual measures

A suite of laboratory-based perceptual measures were used for assessment of noise reduction performance. SRT, which measured speech intelligibility in the presence of background noise, was used in all experiments as the primary means for evaluation. Other aspects of listening in noise that were evaluated were sound quality and acceptable noise level. Additionally, intelligibility in the absence of noise was assessed with a word recognition task.
3.3.1 Word recognition in quiet (CNC)

Speech intelligibility in quiet was assessed by presenting open-set mono-syllabic Consonant-vowel Nucleus-Consonant (CNC) words presented at 60 dB SPL from a loudspeaker located 1.2 m directly in front of the listener (0 degrees) in the absence of any competing noise. The purpose was to check for any undesirable distortion artefacts of the noise reduction processing, which could lead to rejection of the algorithm. The initial consonant, vowel and final consonant were scored phonetically based on the subjects’ response.

CNC word test protocol

Subjects listened to lists of 50 words processed under each of the test conditions. In the first session, the processing condition test order was randomised, and in the second session the test order was reversed providing counter-balanced testing of processing conditions within each subject. A total of two lists (total of 100 words) were presented for each condition. The results were averaged over the two test sessions.

3.3.2 Sentence recognition in noise (SRT)

Speech intelligibility in noise was assessed using the Australian Sentence Test In Noise (AuSTIN, Dawson et al., 2013). The test is an adaptive speech-in-noise test used to find the SRT defined as the SNR for 50% morphemes correct. Short open-set BKB-like sentences spoken by an Australian female were presented at 65 dB SPL in the presence of continuous background noise, the level of which was adapted based on the subjects’ response. According to the AuSTIN, the noise level was increased if more than 50% of the morphemes in the sentence were repeated correctly; otherwise the noise level was reduced. The noise level was adjusted by 4 dB for the first four sentences, and by 2 dB for the remaining 20 sentences. Following each adjustment there was a three second period of noise at the new level before the next sentence was presented. In addition to an adjustment to the noise level, the spatial configuration of the noise sources also changed at that time. These details are described in Section 3.4. The SRT was calculated as the mean SNR for the final 16 sentences. Figure 3-1 shows the software used to measure SRT with an example of how the SNR adapted over the course of an SRT measurement.

SRT test reliability

Due to the varying spatial configuration of the competing noise used for evaluation, it was important to establish the reliability of the SRT test to ensure it was a useful evaluation tool. Test reliability was quantified by analysing data from a series of previous experiments conducted at Cochlear Ltd in 4-TB-rear\(^1\) noise configuration. SRT results were collected under various sound processing conditions and data was analysed when test-retest data was available.

---

\(^1\) This noise type contained four competing talkers that were spatially separated in the rear hemi-field. See Section 3.4 for further details.
Figure 3-1: Example screen shot from the SRT test. The left of this screen shows test configuration settings and the right hand plot shows the adaptation of an SRT run where the SNR adapted up and down during the test based on the subject’s responses.

Figure 3-2: Test reliability in randomly allocated spatially separated 4-talker babble (4-TB-rear)
Table 3-1: Test reliability reported as the standard deviation (SD) in dB between test re-test SRT for various speech in noise tests, for normal hearing (NH) and cochlear implant (CI) subjects

<table>
<thead>
<tr>
<th>Test</th>
<th>SD (NH)</th>
<th>SD (CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HINT (Nilsson et al., 1994)</td>
<td>1.13</td>
<td></td>
</tr>
<tr>
<td>LIST (van Wieringen and Wouters, 2008)</td>
<td>1.17</td>
<td></td>
</tr>
<tr>
<td>VU (Versfeld et al., 2000)</td>
<td>1.07</td>
<td></td>
</tr>
<tr>
<td>BKB-SIN (Etymotic, 2005)</td>
<td>0.8</td>
<td>1.6</td>
</tr>
<tr>
<td>AuSTIN (Dawson et al., 2013)</td>
<td></td>
<td>1.09</td>
</tr>
<tr>
<td>Proposed (4-TB-rear)</td>
<td></td>
<td>1.12</td>
</tr>
</tbody>
</table>

That is, where a subject was tested using a particular processing condition and two SRTs were measured in that condition in the same test session. These SRTs were used as test-retest pairs to measure test reliability by considering the within-subject standard deviation of repeated measurements. The speech processing conditions used were specific to the study which was conducted at the time, and the test-retest reliability was considered only for within speech processing condition.

In total, 252 test-retest data pairs were collected and analysed for 4-TB-rear noise type. The data comprised 30 different subjects and 8 different speech processing conditions. The difference between test and re-test SRT for 4-TB-rear is plotted as a histogram in Figure 3-2.

The test reliability had a standard deviation of 1.12 dB and was in the same order of magnitude as other well-known tests (Table 3-1). In fact, most other tests report test reliability using normal hearing subjects, whereas CI subjects usually show more variability for the same speech material (e.g. BKB-SIN).

This analysis provides evidence that the spatially separated moving noise test environment did not introduce too much variability in the test measurement itself, and is thus useful as a reliable and accurate speech perception test.

SRT test protocol

For all experiments except Experiment V, SRTs for each processing condition were calculated after two SRT runs. The test order of processing conditions was randomised for each subject for the first SRT run, and the test order was reversed for the second SRT run to produce counter-balanced testing (e.g. ABBA or BAAB) for each subject. This test order was designed to minimise systematic learning or training effects involved with listening to the sound processing under evaluation or with the test procedure itself. SRTs were averaged over the two SRT runs.

In Experiment V, one SRT was collected per condition, and therefore no counterbalancing took place; however, the test order was randomised for each subject. Within each test session, a practice SRT was administered at the beginning of each session using a randomly chosen

---

3 There was potentially some influence of sound processing condition on the reliability of the test since algorithm operation could have been affected by the test SNR and changing spatial configuration. The variability due to CI recipients could not be separated from the variability due to the speech processing used to deliver the stimulus and, therefore, the test reliability represents a combination of both effects.
processing condition from those to be evaluated then all processing conditions were subsequently evaluated in a random order.

3.3.3 Sound quality rating (MUSHRA)

Sound quality ratings were performed using a method called ‘Multiple Stimulus with Hidden Reference and Anchor’ (MUSHRA, ITU-R.Recommendation.BS.1534-1, 2003). The test involved judging the sound quality of a number of test conditions as well as a reference condition and an anchor. The reference was the stimulus at +20 dB SNR processed with the Standard condition and, due to the high SNR, was expected to have superior sound quality to all test conditions and hence produce the highest rating. The anchor was the Standard test condition at the test SNR, representing the minimum noise reduction processing of all conditions tested. It was expected to have the worst sound quality, and hence produce the lowest rating, although some noise reduction processing may have made sound quality poorer than the Standard condition. Subjects were presented with a computer touch screen with buttons to play stimuli processed by each of the test conditions (Figure 3-3). One of the buttons was labelled ‘Reference’ and was used to play the reference signal whilst the other buttons were labelled with letters (for example, ‘A’ through ‘F’) and were used to play the test conditions including the anchor and the hidden reference. For each comparison, the allocation of test conditions to the buttons was randomised and hidden from the subject. Pressing the buttons caused the software to replay the same sound sample, changing the underlying sound processing directly. The task involved rating each on a scale labelled ‘very poor’, ‘poor’, ‘fair’, ‘good’, ‘excellent’. Subjects were encouraged to replay the sample and compare ratings between all conditions, including the visible reference. Subjects were instructed that one of the buttons labelled ‘A’ through ‘F’ contained a hidden version of the reference, and when it was detected the subject should mark that sample as ‘excellent’. Subjects were encouraged to use the full scale, but no further requirement was placed on using the rating scale. For example, subjects were not explicitly required to use the ‘very poor’ rating. Sound quality ratings were converted to a numerical score between 0 and 100 based on the marker location.

Instructions given to subjects were adapted from ITU-R.Recommendation.BS.1534-1 (2003) as follows:

*Your task is to provide a judgement of basic audio quality for each of the test signals, A to F. First, listen to the reference signal. Next, listen to the test signals and mark the audio quality using the scale, from Poor to Excellent. One of the test signals should sound the same as the reference. Listen out for it, and when you hear it, you should mark the audio quality as Excellent. Feel free to listen to the test signals and reference signal as many times as you wish, in any order. Compare signals against each other to get the best judgement of audio quality for each test signal. When you have scored all test signals, press Next.*
The stimuli used were Australian female speech extracted as 10-12 s segments from a talking book and subsequently embedded in noise. The speech was embedded in noise at a fixed SNR’s of -5, 0, or +5 dB SNR depending on the experiment. The competing talker locations used in the sound quality comparison did not change as in the SRT testing, but used fixed locations in order to restrict variability in subject responses. Changing locations may have led to conditions being directly compared under different spatial configurations, possibly masking the underlying difference between algorithms. The spatial locations of the noise sources used in the MUSHRA are listed for each noise type in section 3.4.2.

**MUSHRA test protocol**

Subjects performed the MUSHRA comparison for each noise type at each SNR to produce a set of results. The order of trials was randomised for each test set that was administered. During a test session, subjects performed a practice set and a test set, and the results of the practice set were discarded.

### 3.3.4 Acceptable noise level (ANL)

Originally called the tolerated noise level (Nabelek et al., 1991), the ANL (acceptable noise level) test involves adjusting the level of background noise to an acceptable level. To perform the task, the subject first listened to running speech read from a talking audio book, without background noise. The task was to adjust the level of the speech using the touch screen interface (Figure 3-4) until it was a comfortable loudness. This level provided the most comfortable level (MCL). Next, background noise was added, and the subject was asked to adjust the level of the background noise using the touch screen (Figure 3-5) to a maximum tolerable level. The acceptable noise level was then calculated as the difference between the noise level and the MCL, and is actually an SNR, not a level.

Instructions given to subjects were according to those administered by Donaldson et al. (2009) as follows:

*You will listen to a story through a loud speaker. After a few moments, select the loudness of the story that is most comfortable for you, as if listening to a radio. The computer will allow you to make adjustments. First, turn the loudness up until it is too loud and then down until it is too soft. Finally, select the loudness level that is most comfortable for you.*

*You will listen to the same story with background noise of several people talking at the same time. After you have listened to this for a few moments, select the level of background noise that is the most you would be willing to accept or put up with without becoming tense and tired while following the story. First, turn the noise up until it is too loud and then down until the story becomes very clear. Finally, adjust the noise (up and down) to the maximum noise level that you would be willing to put up with for a long time while following the story.*
Methods

Figure 3-3: Example subject input screen used in the MUSHRA sound quality evaluation. The subject was asked to listen to the stimulus processed with each of the different conditions by pressing the buttons labelled ‘Reference’ and ‘A’-‘F’ then rate the overall sound quality of each sample according to the scale from ‘Very Poor’ to ‘Excellent’.

Figure 3-4: User interface used to conduct the acceptable noise level (ANL) test. This screen shows the up/down buttons used to change the level of running speech without background noise present in order to obtain the most comfortable level (MCL).

Figure 3-5: User interface used to conduct the acceptable noise level (ANL) test. This screen shows the up/down buttons used to change the level of the noise to obtain the acceptable noise level.
3.4 Noise Environment

For the experiments in this thesis, four different noise environments were defined with the aim of representing a subset of real-life scenarios that vary in difficulty for the listener. They aim to illustrate important situations in which noise reduction may be of benefit to CI users. The noise environments used in the experiments were designed to improve on traditional tests that have fixed noise locations. The aim was to represent real-life environments where talkers are spatially separated and involved in conversation. To emulate this, the interfering maskers had different locations that varied during the test. The algorithm under evaluation was thus forced to adapt to the varying environment. The test is suitable for the evaluation of a wide variety of noise reduction algorithms, including multi-microphone and binaural algorithms.

Four different noise conditions were defined, and are described in further detail in section 3.4.2

- Two sources of speech weighted noise (2-SWN-rear)
- 4-talker babble rear half (4-TB-rear)
- 4-talker babble full circle (4-TB-full)
- 20-talker babble full circle (20-TB-full)

The environment used for testing in the laboratory was created by presenting sound stimuli to the subject via a set of acoustic loudspeakers arranged in circle. The subject was asked to sit on a chair in the centre of the loudspeaker circle with radius of 1.2 m and the loudspeakers were approximately at ear-height. The loudspeaker circle consisted of 12 loudspeakers, separately equally by 30 degrees.\(^3\)

The different noise conditions differed in the number of noise sources (2, 4, or 20), the type of noise source (SWN or competing speech), and the spatial locations of the noise sources (rear half, or full circle). For the rear half noise types (2-SWN-rear and 4-TB-rear), the spatial locations of the competing sources were restricted to loudspeaker locations in the rear hemi-field.\(^4\) In contrast, for the full circle noise types, the spatial locations of the noise sources were spread around the entire loudspeaker circle,\(^5\) excluding 0 degrees where the target speech was presented from.

3.4.1 Noise sources

The rationale for evaluating in the presence of competing talkers (babble) was to represent real-life situations where competing conversations take place nearby. This would be a common situation in a restaurant or café, for example. Difficulty with communication in this environment can lead to personal isolation from the social situation and was considered an

---

\(^3\) Loudspeakers located at 0, 30, 60, 90, 120, 150, 180, 210, 240, 270, 300, and 330 degrees, relative to the subject’s position at the centre of the circle.

\(^4\) Using loudspeaker locations at 90, 120, 150, 180, 210, 240 and 270 degrees.

\(^5\) Using loudspeaker locations at 30, 60, 90, 120, 150, 180, 210, 240, 270, 300, and 330 degrees.
important situation in which to perform evaluation. Therefore, babble was selected as the predominant noise source for evaluation throughout all experiments. The 4-talker babble situations represent a small café or restaurant, whilst 20-talker babble is more representative of a large social gathering or reception environment. In the 2-noise source scenario (2-SWN-rear), competing talkers were replaced by SWN noise sources. This was to enable CI users to easily identify the target speech from the background noise\textsuperscript{6}. Another advantage of using SWN was that the test suite included a condition with non-modulated noise, which is a noise condition that is particularly suited to single channel noise reduction techniques. One of these was evaluated in Experiments II and V (Chapters 5 and 8). In modulated noise (babble), single channel noise reduction techniques do not generally provide benefit.

The SWN sources were generated from two uncorrelated white Gaussian noise sources that were filtered to have the same spectrum as that of the target sentence material used in the SRT test. This was equal to the international long-term average speech spectrum (ILTASS) defined by Byrne et al. (1994). The competing babble was generated from four independent recordings made separately of two males and two females reading aloud from a newspaper for approximately 3 minutes. The recordings were made in a sound treated audiometric booth at the Bionic Ear Institute, Melbourne. Each recording was adjusted to have a long-term spectrum equal to the ITLASS.

### 3.4.2 Spatial configuration

During a typical situation that contains competing speech from a nearby conversation, the location of competing speech may change location as the conversation shifts from person to person. This effect was represented in the test environments by changing the location of competing talkers during the test through a randomisation procedure.

To create the 2-SWN-rear environment, two uncorrelated sources of SWN were assigned to random locations drawn from the set of seven locations in the rear half of the circle (from 90 to 270 inclusive, separated by 30 degrees). Randomisation of the SWN locations was assigned according to a uniform distribution at the start of each sentence during the test, and allocation of both SWN sources to the same location was allowed, as generated by the randomisation.

To create the 4-TB-rear environment, four independent talkers were assigned random locations drawn from the set of seven locations in the rear half of the circle (from 90 to 270 inclusive, separated by 30 degrees). Randomisation of talker locations was assigned according to a uniform distribution at the start of each sentence during the SRT test, and allocation of multiple talkers to any location was allowed. The locations were restricted to the rear half to emulate a real-life situation with four competing talkers, where the listener might be able to orient themselves within the room to ensure the relatively small number of competing talkers to the side and/or rear of the listener.

---

\textsuperscript{6}Typically, speech tests that contain only one or two competing talkers utilise talker gender as a way to identify the target speech, for example, target female speech in the presence of male competing talkers. Normal hearing listeners have some ability to separate the two, largely based on voice pitch, but this is a difficult task for CI users.
To create the 4-TB-full environment, the four talkers were spread around the full circle as opposed to restricting them to the rear half, reflecting an environment where the listener may not be able to orient the noise sources behind them. Each of the four talkers was assigned to one of the four quadrants of the loudspeaker circle to ensure a spread of competing noise around the full circle. The location of each talker was randomised amongst two locations within each quadrant. In the first quadrant, the location was randomly chosen from 30 and 60 degrees, in the second quadrant from 120 and 150, in the third quadrant from 210 and 240, and in the fourth quadrant from 300 and 330. The competing talkers were allocated to one quadrant and the randomisation procedure did not change talkers from one quadrant to another. In this way, the talker-to-quadrant allocation remained fixed but the location within the quadrant was randomised.

To create the 20-TB-full environment, each of the four independent talker recordings was divided into five uncorrelated segments to produce 20 different speech segments in total. These 20 segments were assigned random locations drawn from the set of 11 locations around the full circle excluding 0 degrees (from 30 to 330 inclusive, separated by 30 degrees). In a real-life situation with 20 competing talkers, it is unlikely a listener would be able to orient themselves with all competing talkers to the side and/or rear. Therefore, the 20-talker babble was presented from the full circle (excluding 0 degrees) during the test in order to more closely represent a situation with relatively many competing talkers.

For all four noise types, the timing of the changes to the randomisation was dependant on the test being executed. The SRT test consisted of presenting short sentences and adapting the level of the noise after each sentence based on the subjects’ response. For the SRT test, the spatial randomisation was performed for each sentence and the noise was presented in that configuration for the entire sentence duration, including three seconds before the sentence actually commenced. The three second lead-in allowed some time for the algorithm under test to stabilise, both to the new noise level due to the adaptive test and to the new spatial configuration, before the sentence began.

The MUSHRA and ANL test both involved listening to a recorded book, and were made over a relatively short duration of stimulus, in the order of 10 seconds or so. It was not desirable to change the spatial configuration whilst switching programs, and not possible to involve many samples of the spatial randomisation in these tests. A single spatial configuration was chosen to allow for consistent comparisons amongst different programs. Without such a control, different randomisations for different conditions may have led to comparisons between conditions being undertaken with different spatial configurations, potentially masking differences between conditions.

The theoretical measurement using STI and error rates used a specific randomisation schedule that was repeatable. The sequence involved a 30 second schedule where the randomisation changed every three seconds.

A summary of the specific randomisations, spatial configurations and test usage are detailed in the following sections for each noise type.
Summary of 2-SWN-rear configuration

Noise sources: Two sources of uncorrelated speech-shaped noise
Spatial configuration: Rear hemi-field (Figure 3-6)
Randomisation (Table 3-2): SRT – all sources randomised for each sentence
                       MUSHRA/ANL – fixed
                       STI/Error rates – 1 of 2 noise source randomised every 3s
Wav filename: ILTASS2rH

Table 3-2: Spatial configurations for 2-SWN-rear noise type showing number of noise sources at each spatial location for the different tests. (* indicates randomly chosen locations)

<table>
<thead>
<tr>
<th>Test</th>
<th>Time point</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>30</td>
<td>60</td>
</tr>
<tr>
<td>SRT Randomised each sentence</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>MUSHRA/ANL Fixed</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>STI/Error Rates 0-3 s</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3-6 s</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>6-9 s</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>9-12 s</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>12-15 s</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>15-18 s</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>18-21 s</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>21-24 s</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>24-27 s</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>27-30 s</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Figure 3-6: Loudspeaker locations used for 2-SWN-rear noise configuration
Summary of 4-TB-rear configuration

Noise sources: Four sources of competing speech (2 male, 2 female)
Spatial configuration: Rear hemi-field (Figure 3-7)
Randomisation (Table 3-3): SRT – all sources randomised for each sentence
MUSHRA/ANL – fixed
STI/Error rates – 1 of 4 source randomised every 3s
Wav filename: BEIBABB4rH

Table 3-3: Spatial configurations for 4-TB-rear noise type showing number of noise sources at each spatial location for the different tests. (*) indicates randomly chosen locations

<table>
<thead>
<tr>
<th>Test</th>
<th>Time point</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>30 60 90 120 150 180 210 240 270 300 330</td>
<td></td>
</tr>
<tr>
<td>SRT Randomised each sentence</td>
<td>- - * * * * * * * *</td>
<td></td>
</tr>
<tr>
<td>MUSHRA/ANL Fixed</td>
<td>- - - 1 1 - 1 1 - - -</td>
<td></td>
</tr>
<tr>
<td>STI/Error Rates 0-3 s</td>
<td>- - - 1 - 2 - 1 - - -</td>
<td></td>
</tr>
<tr>
<td>3-6 s</td>
<td>- - - 1 - 1 1 1 - - -</td>
<td></td>
</tr>
<tr>
<td>6-9 s</td>
<td>- - - - - 1 1 1 1 - -</td>
<td></td>
</tr>
<tr>
<td>9-12 s</td>
<td>- - - - - 1 1 - 2 - -</td>
<td></td>
</tr>
<tr>
<td>12-15 s</td>
<td>- - - - - 1 1 - 2 - -</td>
<td></td>
</tr>
<tr>
<td>15-18 s</td>
<td>- - - 1 - 1 - - 2 - -</td>
<td></td>
</tr>
<tr>
<td>18-21 s</td>
<td>- - - 1 1 1 - - 1 - -</td>
<td></td>
</tr>
<tr>
<td>21-24 s</td>
<td>- - 1 - 1 1 - - 1 - -</td>
<td></td>
</tr>
<tr>
<td>24-27 s</td>
<td>- - 1 1 1 - - 1 - -</td>
<td></td>
</tr>
<tr>
<td>27-30 s</td>
<td>- - 2 1 - - - - 1 - -</td>
<td></td>
</tr>
</tbody>
</table>

Figure 3-7: Loudspeaker locations used for 4-TB-rear noise configuration
Summary of 4-TB-full configuration

Noise sources: Four sources of competing speech (2 male, 2 female)
Spatial configuration: Full circle restricted (Figure 3-8)
Randomisation (Table 3-4): SRT – all sources randomised for each sentence
                          MUSHRA/ANL – fixed
                          STI/Error rates – 4 of 4 noise sources randomised every 3s
Wav filename: BEIBABB4prF

Table 3-4: Spatial configurations for 4-TB-full noise type showing number of noise sources at each spatial location for the different tests. (*) indicates randomly chosen locations

<table>
<thead>
<tr>
<th>Test</th>
<th>Time point</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>30</td>
<td>60</td>
</tr>
<tr>
<td>SRT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Randomised</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>each sentence</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MUSHRA/ANL</td>
<td>Fixed</td>
<td>1</td>
</tr>
<tr>
<td>STI/Error Rates</td>
<td>0-3 s</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>3-6 s</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>6-9 s</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>9-12 s</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>12-15 s</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>15-18 s</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>18-21 s</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>21-24 s</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>24-27 s</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>27-30 s</td>
<td>-</td>
</tr>
</tbody>
</table>

Figure 3-8: Loudspeaker locations used for 4-TB-full noise configuration
3.4 Noise Environment

Summary of 20-TB-full configuration

Noise sources: Five uncorrelated segments from each of four sources of competing speech (2 male, 2 female), totalling 20 competing speech sources

Spatial configuration: Full circle (Figure 3-9)

Randomisation (Table 3-5):
- SRT – all sources randomised for each sentence
- MUSHRA/ANL – fixed
- STI/Error rates – 5 noise sources randomised every 3s

Wav filename: BEIBABB20rF

Table 3-5: Spatial configurations for 20-TB-full noise type showing number of noise sources at each spatial location for the different tests. (* indicates randomly chosen locations)

<table>
<thead>
<tr>
<th>Test Type</th>
<th>Time point</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRT</td>
<td>30 60 90 120 150 180 210 240 270 300 330</td>
<td></td>
</tr>
<tr>
<td>Randomised each sentence</td>
<td>* * * * * * * * * *</td>
<td></td>
</tr>
<tr>
<td>MUSHRA/ANL</td>
<td>0-3 s</td>
<td>2 2 3 1 2 2 4 1 2 2 1</td>
</tr>
<tr>
<td>STI/Error Rates</td>
<td>3-6 s</td>
<td>1 3 1 3 4 1 1 3 2 1</td>
</tr>
<tr>
<td></td>
<td>6-9 s</td>
<td>1 3 2 1 4 3 1 2 2 1</td>
</tr>
<tr>
<td></td>
<td>9-12 s</td>
<td>1 4 - 2 4 2 1 1 3 2</td>
</tr>
<tr>
<td></td>
<td>12-15 s</td>
<td>3 4 - 1 4 3 1 1 2 1</td>
</tr>
<tr>
<td></td>
<td>15-18 s</td>
<td>1 3 2 2 1 3 1 2 - 2 3</td>
</tr>
<tr>
<td></td>
<td>18-21 s</td>
<td>1 2 3 2 - 4 2 1 - 3 2</td>
</tr>
<tr>
<td></td>
<td>21-24 s</td>
<td>1 2 4 2 - 5 2 1 1 1 1</td>
</tr>
<tr>
<td></td>
<td>24-27 s</td>
<td>1 1 4 - 1 3 1 3 2 3 1</td>
</tr>
<tr>
<td></td>
<td>27-30 s</td>
<td>1 1 3 - - 4 2 5 1 2 1</td>
</tr>
</tbody>
</table>

Figure 3-9: Loudspeaker locations used for 20-TB-full noise configuration
3.4.3  Sound level and calibration

Target speech was always presented from the loudspeaker at 0 degrees directly in front of the listener at a presentation level of 65 dB SPL. The level was calibrated by presenting a 1 kHz narrow band noise stimulus and measuring with a sound level meter with the microphone located at the centre of the loudspeaker circle, without the subject or manikin in place. The microphone was attached to a long cable which hung from the ceiling so that the sound level meter was not inside the loudspeaker circle at the time of measurement. The narrow band noise had a digital level as stored in the wave file of -25 dB FS RMS. This was the same digital level used for all target speech and noise recordings. The level of the 1 kHz narrow band noise was adjusted until the recorded level was 65 dB SPL, measured with A-weighting and averaged over at least three seconds of stimulus. Amplification headroom was allowed so that the maximum level the system could achieve was 100 dB through a single loudspeaker. All sound material was presented with the same digital level as the calibration sound produced the same SPL as the calibration sound. Each loudspeaker around the circle was calibrated in the same fashion, and the level adjustment for each was saved in a calibration file for the sound card. The settings in the file were activated at the beginning of each test session.

The reported noise level in all cases was the total noise level, that is, the sound pressure level of the combined sum of all noise sources. Therefore, given a specified noise level, the level of an individual competing talker was lower in the 20-TB noise compared to an individual competing talker in the 4-TB noise.

3.4.4  Stimulus presentation

Depending on the experiment, two different methods of sound presentation were available. The first method used direct acoustic stimulation via loudspeakers with the subject in the test room. The second method used an indirect method of recording microphone impulse responses allowing offline reproduction of the microphone signals that were ultimately delivered to the recipient via audio cable. For the second method, subjects were not required to physically visit the test room where the impulse responses were recorded.

To record the impulse responses, a Brüel & Kjær head and torso simulator (HATS) was placed in the middle of the test room. A CS2452 Response loudspeaker was used to generate a swept sine stimulus for measuring the impulse response between the loudspeaker and the microphones located on the ears of the HATS. The loudspeaker location was moved to record an impulse response at 15 degree intervals around the circle, at distances of 1, 2 and 3 m from the listening position, although not all of these positions were used in the evaluation. The device used was a modified version of the CP810 BTE sound processor which had only the microphones and a cable connection to an external preamplifier. The two microphones were mounted in the BTE shell in the same positions as the commercial device where the separation is approximately 12 mm. Recordings were made bilaterally with two different devices so a total of four microphone signals were recorded simultaneously. However, only the left ear data was used in the evaluation. An RME Multiface II was used under MATLAB control for playback and recording of the swept sine stimulus at a sampling frequency of 44.1 kHz. The frequency response of the loudspeaker was measured under anechoic conditions, and was
used to equalise the frequency response so that the effect of the loudspeaker was removed from the recorded impulse responses.

For some conditions in Experiment III, the impulse response method was also used for the sound treated room. However, there were some subtle differences in the materials and methods used to record these: a Knowles KEMAR manikin was used as opposed to HATS, the recording distance was 1.2 m, a maximum length sequence (MLS) technique was used to obtain the impulse response, and the response of the loudspeaker was not removed because the impulse response at each location was recorded with different loudspeakers (of the same model, Genelec 8020a). These loudspeakers had all been previously calibrated to produce the same sound level at 1 kHz at the listening position without a human test subject or KEMAR manikin present.

The impulse responses recorded in both rooms were calibrated at each distance by producing a 1 kHz tone at an angle of 0 degrees. The output was measured and averaged across the four recording microphones. All impulse responses in the set recorded at the same listening distance were amplified by the same factor, so the response at 0 degrees was -25 dB FS. The impulse responses at each of the other distances were calibrated similarly so that an input signal at 0 degrees from any distance (in either test room) produced the same level at the output of the BTE microphones.

The impulse responses were used to generate sound inputs used in the evaluation by convolving the impulse response with the desired stimulus at a sampling frequency of 44.1 kHz. The convolved microphone signals were routed via a sound card and cabled directly to the input of the computer-based real-time processing system. The direct input signal levels were further adjusted so that the software producing the stimulus had a reference to the sound pressure level at the microphone input. In this way, all stimuli were calibrated to a sound pressure level at the BTE microphone input with the manikin head in position.

As a consequence, there was an important, but necessary difference in the calibration procedure used for sound presented live compared to sounds generated using impulse responses. In the live presentation, the calibration position was at the centre of the loudspeaker circle without the manikin in position. In the offline presentation of sounds generated via impulse response, the calibration position was at the ear-level microphone with the manikin in position at the centre of the circle. Given this difference, comparison of absolute signal level between the two situations is not directly possible. This is due to the acoustic effects of the manikin in position, and the difference in location of the microphones used in the calibration procedure.
Figure 3-10: The sound treated Whitford booth located at Cochlear, East Melbourne, used for speech tests and evaluation. Reverberation time $T_{30} = 87$ ms.

Figure 3-11: Reverberant training room used for evaluation of the noise reduction algorithms. Reverberation time $T_{30} = 523$ ms.
3.5 Test rooms

Two different rooms were used to present acoustic material during the tests. One room was a sound treated audiometric room located at Cochlear Ltd., East Melbourne, known as the Whitford Booth (Figure 3-10). The room had dimensions 4 m x 5 m x 2 m (L x W x H), the floor was carpeted and both the walls and ceiling were treated with acoustic panels to produce relatively low reverberation ($T_{30} = 87$ ms).

The other room that was utilised for evaluation was not designed for audiometric testing, but as a meeting and training room, and was located at the Australian Hearing head office in Greville St, Chatswood, Sydney Australia, known as the “Training Room” (Figure 3-11). The dimensions of the room were 11.8 m x 8.6 m x 3.6 m (L x W x H). Two adjoining walls were painted concrete brick surfaces, lightly covered with posters and pictures, with two wooden doors. Another wall was covered with plaster and the remaining wall was almost entirely glass. The ceiling of the room was treated with suspended absorbing panels and the floor was covered with carpet. The reverberation time was 523 ms.

3.5.1 Reverberant properties of the rooms

The reverberant properties of the rooms, including the reverberation time, were estimated from the energy decay curve obtained from the time-reversed integral of the impulse response (Schroeder, 1965). The decay time ($T_{30}$) is an estimation of room reverberation time ($RT_{60}$), the time at which the reverberant energy has decayed to 60 dB below the level of the direct sound (Beranek, 1988). $T_{30}$ was estimated by fitting a straight line to the energy decay curve over the range -5 to -35 dB and extrapolating to the -60 dB point. The early decay time (EDT) was also estimated and is known to relate to the perceived reverberation of the room. It was calculated by fitting a straight line to the energy decay curve over the range -1 to -10 dB and extrapolating to the -60 dB point. The direct-to-reverberant (DRR) and clarity ($C_{50}$ and $C_{7}$) are measures that assess the strength of the direct sound field and are used to estimate speech intelligibility in reverberant spaces. Low values are representative of high reverberation. They were estimated by calculating the ratio of energy in the direct portion of the impulse response to the reverberant part of the impulse response. The time point used to split the direct from reverberant portion was 50 ms ($C_{50}$), 7 ms ($C_{7}$), and 2.5 ms (DRR) measured from the onset of the impulse response. The onset of the impulse response has been defined as the time point where the energy in the impulse response is 20 dB greater than the noise floor (Zahorik, 2002), however a threshold of 40 dB was used to avoid errors in automatic onset detection. The $T_{30}$, EDT, $C_{50}$, $C_{7}$ and DRR were estimated for each of the four microphones at 12 angles (from 0 to 330 in 30 degree steps) to form a set of 48 values. The estimation was calculated as the median across the set of values, and estimated separately at each listening distance (Table 3-6).
Table 3-6: Room reverberation properties for the rooms and listening distances used in the evaluation. Reverberation time (T30), early decay time (EDT), clarity indices (C50, C7) and direct-to-reverberant ratio (DRR) have been estimated from the impulse response data at each listening distance.

<table>
<thead>
<tr>
<th>Test room</th>
<th>Distance (m)</th>
<th>T30 (ms)</th>
<th>EDT (ms)</th>
<th>C50 (dB)</th>
<th>C7 (dB)</th>
<th>DRR (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whitford booth</td>
<td>1.2</td>
<td>87</td>
<td>7</td>
<td>35.2</td>
<td>15.3</td>
<td>10.8</td>
</tr>
<tr>
<td>Training room</td>
<td>1</td>
<td>460</td>
<td>20</td>
<td>17.2</td>
<td>12.0</td>
<td>11.1</td>
</tr>
<tr>
<td>Training room</td>
<td>2</td>
<td>528</td>
<td>402</td>
<td>11.8</td>
<td>5.9</td>
<td>4.5</td>
</tr>
<tr>
<td>Training room</td>
<td>3</td>
<td>581</td>
<td>412</td>
<td>9.2</td>
<td>3.3</td>
<td>-0.3</td>
</tr>
</tbody>
</table>

3.5.2 Room simulator

A room simulator (MCRoomSim) was utilised for Experiment III (Chapter 6). It allowed the bilateral CP810 microphones to be modelled inside a simulated shoe-box environment with arbitrary acoustic properties, defined for each wall, floor and ceiling. The simulator was utilised to produce microphone signals that were more extreme in reverberation than the Training room. Further details regarding the configuration and validation of the simulator are provided in Experiment III.

3.6 Field-based outcome measures

Field-based outcomes measures were used in the final experiment (Chapter 8) to evaluate user preference away from the laboratory. A traditional paper-based questionnaire was administered and subjects undertook a real-time voting task. The voting task required users to change listening programs and vote for their preferred setting using their sound processor’s remote control.

3.6.1 Questionnaire

The SSQ questionnaire (Gatehouse and Noble, 2004) was adapted to acutely compare listening programs that were trialled during the take-home component of the evaluation in Experiment V (Chapter 8). The original SSQ had 49 questions in total, divided into three sections; Speech, Spatial and Quality, which were further divided into pragmatic subscales (Gatehouse and Akeroyd, 2006). For the purpose of this study, the Speech section was used which contained 14 questions in total, divided into four pragmatic subscales according to Gatehouse and Akeroyd (2006): Speech in quiet (items 2 & 3), Speech in noise (items 1, 4, 5 & 6), Speech in speech contexts (items 7, 8, 9 & 11) and Multiple stream-speech processing and switching (10, 12 & 14). The SSQ response sheet was modified to provide two rating scales for each of the 14 questions, so that P1 and P2 could be rated on separate scales for each question (an example is shown in Figure 3-12).

7 version 5.6
“You are talking with one other person and there is a TV on in the same room. Without turning the TV down, can you follow what the person you’re talking to says?”

![Figure 3-12: Example response sheet from the modified SSQ questionnaire that allowed simultaneous ratings of two listening programs, not possible with the original SSQ format.](image)

Further questions regarding overall program preference were also separately posed for quiet and noisy situations. For those additional two questions, subjects were asked to rate their overall preference along a seven point scale that was marked as “Much prefer P1”, “Moderately prefer P1”, “Slightly prefer P1”, “No difference between P1 and P2”, “Slightly prefer P2”, “Moderately prefer P2”, and “Much prefer P2”. The rating was converted to a seven point numerical scale from -3 to +3 for the purpose of analysis. Subjects were also invited to make open comments regarding their listening experience with the two programs. They were also invited to comment on the voting experience itself. The full questionnaire that was administered can be found in Appendix A.

### 3.6.2 Voting

One of the main objectives of Experiment V (Chapter 8), which utilised the CP900 sound processor, was to obtain subjective feedback from subjects after using the SpatialNR algorithm in their everyday lives. This was achieved by devising a voting paradigm based on the work by Zakis et al. (2007) that enabled subjects to identify their preferred listening program using their processor’s remote control device (CR230). Subjects were able to vote at any time, so were not restricted to evaluation under specific acoustic situations, nor were they required to memorise the acoustic situation and preferred listening program to complete this task. In fact, using the data logging features of the accompanying CR230 device and the sound classification feature of the CP900, both the acoustic scene and preferred program selection were obtained directly in the listening environment.
Table 3-7: CP900 SCAN sound classes and associated directionality that is automatically selected

<table>
<thead>
<tr>
<th>Sound Class</th>
<th>CP900 directionality selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quiet</td>
<td>Standard</td>
</tr>
<tr>
<td>Speech</td>
<td>Standard</td>
</tr>
<tr>
<td>Speech in Noise</td>
<td>Beam</td>
</tr>
<tr>
<td>Noise</td>
<td>Zoom</td>
</tr>
<tr>
<td>Music</td>
<td>Standard</td>
</tr>
<tr>
<td>Wind</td>
<td>Wind directionality</td>
</tr>
</tbody>
</table>

CP900 sound classification

The CP900 sound processor has a signal processing feature called SCAN, which classifies the acoustic scene and makes changes to the sound processing under automatic control (Mauger et al., 2014). The six sound classes are Quiet, Speech, Speech in Noise, Noise, Music, and Wind, shown in Table 3-7. Classification is based on extracting acoustic features from the incoming signal, and then using a decision tree to determine the sound class. The classifier is trained on a dataset of manually labelled acoustic recordings. The classifier is used to automatically control the directionality setting of the CP900 according to Table 3-7.

CR230 remote control

The CR230 remote control is a handheld remote control device for the CP900 sound processor that allows user control over various sound processor settings. Some of these settings are the listening program, volume, sensitivity, and telecoil amongst others. There are also a range of other control and monitoring features including display of the current sound class (Mauger et al., 2014). The CR230 also has an extensive data logging feature that records the time and date of a range of events that occur both on the sound processor and on the CR230 itself. For example, the CR230 can log all user interaction with the device including all user button presses, as well as some events that occur on the CP900 BTE such as listening program changes and scene classification changes.

Voting paradigm

The data logging capability led to the design of a voting paradigm using the CR230. The button on the side of the CR230, at the upper left, is normally used to enable the telecoil feature in the CP900 sound processor. However, for the purpose of voting, the telecoil functionality of the sound processor was disabled, and therefore pressing the button on the CR230 had no effect. However, the action of the button press was still logged by the CR230 data logging feature, and was thus used and interpreted as a vote event.
3.6 Field-based outcome measures

The other key components of data logging were the scene classification changes and listening program changes that were logged by the CR230. Consequently, the data logs of vote events, scene classification change events and listening program change events provided time series data that could be analysed to provide direct information about preferred listening program in a range of different acoustic environments, as determined by the scene classification. Figure 3-13 shows the CR230 remote control and CP900 sound processor used for voting experiments.

Subjects were issued with a sound processor that had two (or three) listening programs that had the telecoil functionality disabled. Subjects were asked to compare the programs in as many different acoustic environments as possible, and vote for their preferred program using the CR230. At the completion of the two week period, the data log was retrieved from the CR230 and analysed.

Specific instructions provided to subjects for the voting task were;

We have changed your settings so that your telecoil is now disabled. The telecoil button on your remote control will be used for another purpose. Here’s how it works;

You have 2 (or 3) new programs in your sound processor. Your task over the next week is to compare them in as many situations as possible. In each situation, compare the programs in any order and as many times as you wish. Any time you find that one program is more helpful than the other(s), vote for it by pressing the telecoil button when you are listening to the most helpful program. You may find that the most helpful program is different in different situations. In this case, vote for the program you find most helpful at the time. You are encouraged to be a swinging voter if needed. In some situations, you may find that there is no difference between programs, and that no program is more helpful than the other(s).
In this case, sit on the fence and don’t vote, simply don’t press the telecoil button. If you accidentally vote for the wrong program, don’t worry, change programs and vote for the correct one. Over the next week, vote as many times as you can, by comparing programs and voting for the most helpful. There is no limit to the number of times you can vote.

Voting data analysis

The raw voting data consisted of a series of time-stamped events logged in the CR230. Program changes were logged each time the user changed listening programs. Scene changes were logged each time the SCAN program detected a new scene, and vote events were logged each time the user pressed the vote button. For each vote, the relevant program and scene were determined as follows:

The program choice was determined by the program that was selected at the time the vote was recorded. Two or more consecutive votes where the user did not change listening programs between votes were counted as a single vote. This was done in order to exclude accidental voting (by button presses in the pocket) and to exclude votes that did not coincide with trying the other program(s).

In order to determine the sound class associated with each vote, the detected sound class was analysed over 10 seconds preceding the vote event since the evaluation of programs would likely have occurred over a period of time, possibly under different scene classifications. In cases where the sound class was variable, the vote was assigned according to the dominant sound class over the 10 seconds preceding the vote event, and in the case of an equal distribution, to the most recently detected sound class.

For each data log, the votes were summed and represented in a matrix of program x scene, \( V(p,s) \), where each element represented the number of votes. These data were then converted to a proportion of the total number of votes for each scene,

\[
P(p, s) = \frac{V(p,s)}{\sum_{i=1}^{V(s)} V(i,s)}
\]  \hspace{1cm} (3-9)

where \( p \) indicates the program and \( s \) indicates the scene.

For comparisons that involved only two programs, a normalised vote score was calculated by scaling the proportion of votes that were favour of the experimental program to cover the range -1 to +1,

\[
N_e(s) = 2P(e, s) - 1
\]  \hspace{1cm} (3-10)

where \( e \) indicates the experimental program.

In this way, the normalised vote score was a measure of the central tendency of vote events for each scene from the perspective of the experimental program. A score of +1 represented all votes in favour of the experimental program, -1 represented all votes in favour of the baseline program, and a score of 0 represented equal votes for both programs. The normalised
vote score was calculated to facilitate summarising data across the group, such that subject’s responses contributed equally to the group result, and contributions were not influenced by the number of votes recorded by each subject. That is, the number of votes was not used to weight the normalised vote scores when summarising the group’s data. In addition, to avoid sensitivity to isolated votes from some subjects in some scenes, a minimum of three votes was required for a data point to contribute to the group results for that scene.

3.7 Sound processing equipment

The equipment and devices used in the experiments to perform real-time sound processing of the algorithms under evaluation are described in the following sections.

3.7.1 Simulink and xPC

For all experiments except Experiment V (Chapter 8) which utilised the CP900 sound processor, signal processing was performed on a real-time computer-based speech processor, previously used for similar investigations (e.g. Dawson et al., 2011). The system performed the same processing as the CP810/CP900 commercial device, but had the experimental SpatialNR algorithm integrated for the purpose of evaluation. The model was implemented in Simulink using Mathworks Matlab version R2009a, and real-time processing utilised the xPC Target product. This allowed the model to be constructed graphically using functional processing blocks, which was then compiled and downloaded onto the real time processing PC. The system allowed control and configuration of the real-time model running on the PC via the host PC running the Simulink model. The system used a sampling rate of 15.7 kHz and 128 point FFT frames were acquired every 16 samples at a rate of 489 Hz according to the commercial sound processors. Software control was used to switch between algorithms and to control the SpatialNR parameters under investigation. Parameters for Zoom and Beam conditions were used according to the commercial sound processors and were not adjusted during the experiments. During evaluation, the acoustic signal was captured using a modified CP810 sound processor, which contained only the microphone components that was connected to the real-time processor. The real-time computer based system also provided means for microphone signal generated using the impulse response.

3.7.2 CP900 sound processor

In Experiment V, the SpatialNR algorithm was implemented in experimental firmware for the CP900 BTE sound processor, enabling subjects to wear the device out of the laboratory. The firmware implementation was provided by Cochlear DSP Team, Sydney.

3.7.3 Slow and fast switching

In order to conduct clinical trials comparing various signal processing configurations, it was necessary to switch between conditions. During SRT testing, the speed of switching did not
need to be instantaneous, since each condition was tested sequentially with a break in between each SRT run. However, during the MUSHRA sound quality comparison, there was a need to switch quickly between conditions so that the user could quickly listen to processing under one condition, then another, in order to make direct comparisons. For this reason, switching between conditions needed to be fast, ideally within a few hundred milliseconds, and desirably less than 1 second. Therefore, on the real time system and the CP900, a method of fast switching was introduced to complement the method of slower switching used in SRT testing.

The slow switching method was based on bulk parameter changes, the collection of which is referred to as a map. On both the real time system and CP900, this was achieved by storing map settings in a file, then recalling and loading the map at the appropriate time. The switching speed for a map change on the real time system was approximately 3 seconds, while on the CP900 it was approximately 1 second. This was suitable for SRT testing, but the speed was considered too slow for MUSHRA sound quality comparisons.

Fast switching was based on direct modification of only the parameter values required, as opposed to a bulk change with a map. This was achieved via direct access to parameter values on the real-time execution PC, and via direct access to memory locations on the CP900 sound processor. Parameter changes via this method were essentially instantaneous, making it possible to conduct the MUSHRA test on both platforms. In order to verify that map changes were functionally equivalent to the fast switching method, theoretical measurements were conducted and compared for expected results. This testing ensured that map based switching was equivalent to fast switching, and that the switching between processing conditions produced the expected theoretical measurement. This evaluation also provided a thorough analysis of the entire test equipment/method prior to human evaluation.

3.8 Subjects

Subjects were recruited through the Cochlear Implant Clinic at the Royal Victorian Eye and Ear Hospital, East Melbourne, Australia. All were adults (over 18 years of age) at the time of testing. All subjects volunteered their time and were not paid for their participation, although travel costs to and from the testing laboratory were reimbursed. Subjects visited the test laboratory at Cochlear Ltd, East Melbourne for all testing. Ethics approval was obtained via the Royal Victorian Eye and Ear Hospital human ethics committee and subjects signed a written consent form at beginning of each experiment.

All subjects were implanted with a Cochlear Nucleus implant system and were current users of the CP810 or Freedom sound processor. In the evaluation, users’ existing map parameters were transferred from their own device to the real-time evaluation system. All subjects used the Advanced Combination Encoder\(^8\) (ACE, McDermott et al., 1992; Vandali et al., 2000) program with sound processing feature Autosensitivity (Patrick et al., 2006) enabled. Signal processing features and parameters for Adaptive Dynamic Range Optimization (ADRO\(^8\),

---

\(^8\) One exception was S31 in Experiment IV who used MP3000 (Buechner et al., 2011) program instead of ACE.
Blamey, 2005), C-SPL, stimulation rate and number of maxima are shown for each subject in Table 3-9. For all laboratory-based testing, all subjects were tested monaurally using the real-time system or CP900 BTE to activate the CI in their best performing ear, while any direct acoustic transmission was attenuated using ear-plugs when necessary.

A total of 33 subjects participated across five experiments and some subjects participated in more than one experiment. There were 12, 11, 8, 13, and 15 subjects who participated in experiments I-V, respectively. Specific details concerning individual subject participation in each experiment are provided in Table 3-9.

### 3.9 Experiment summary by outcome measure and noise type

Table 3-8 outlines the laboratory-based perceptual outcome measures that were used in each experiment, including the noise type used for the evaluation. The experiment number (I-V) is used in the table to indicate when each test/noise combination was used.

#### Table 3-8: Summary of outcome measures and noise types used in each of the five experiments

<table>
<thead>
<tr>
<th>Perceptual outcome measure</th>
<th>Silent</th>
<th>2-SRN-rear</th>
<th>4-TB-rear</th>
<th>4-TB-full</th>
<th>20-TB-full</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNC Words</td>
<td>II</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speech reception threshold (SRT)</td>
<td>I</td>
<td>II</td>
<td>III</td>
<td>IV</td>
<td>V</td>
</tr>
<tr>
<td>Sound quality (MUSHRA)</td>
<td>II</td>
<td>II</td>
<td>IV</td>
<td></td>
<td>V</td>
</tr>
<tr>
<td>Acceptable noise level (ANL)</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>V</td>
</tr>
</tbody>
</table>
Table 3-9: Demographic details and cochlear implant map configuration for the subjects who participated in the experiments. R – right ear, L – left ear, HA – hearing aid.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Age (yrs)</th>
<th>Implant use (yrs)</th>
<th>Stimulation rate (Hz)</th>
<th>Number of maxima</th>
<th>Test ear</th>
<th>Non-test ear device</th>
<th>C-SPL</th>
<th>ADRO</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>53</td>
<td>3</td>
<td>900</td>
<td>8</td>
<td>R</td>
<td>HA</td>
<td>75</td>
<td>✓</td>
</tr>
<tr>
<td>S2</td>
<td>73</td>
<td>2</td>
<td>900</td>
<td>12</td>
<td>L</td>
<td>HA</td>
<td>75</td>
<td>✓</td>
</tr>
<tr>
<td>S3</td>
<td>72</td>
<td>3</td>
<td>1200</td>
<td>8</td>
<td>L</td>
<td>Freedom</td>
<td>75</td>
<td>✓</td>
</tr>
<tr>
<td>S4</td>
<td>45</td>
<td>5</td>
<td>1800</td>
<td>8</td>
<td>L</td>
<td>HA</td>
<td>75</td>
<td>✓</td>
</tr>
<tr>
<td>S5</td>
<td>42</td>
<td>3</td>
<td>900</td>
<td>8</td>
<td>L</td>
<td>-</td>
<td>75</td>
<td>✓</td>
</tr>
<tr>
<td>S6</td>
<td>82</td>
<td>2</td>
<td>900</td>
<td>8</td>
<td>R</td>
<td>CP810</td>
<td>75</td>
<td>✓</td>
</tr>
<tr>
<td>S7</td>
<td>65</td>
<td>3</td>
<td>900</td>
<td>8</td>
<td>L</td>
<td>HA</td>
<td>75</td>
<td>✓</td>
</tr>
<tr>
<td>S8</td>
<td>67</td>
<td>3</td>
<td>900</td>
<td>8</td>
<td>R</td>
<td>HA</td>
<td>75</td>
<td>✓</td>
</tr>
<tr>
<td>S9</td>
<td>68</td>
<td>2</td>
<td>500</td>
<td>8</td>
<td>L</td>
<td>HA</td>
<td>75</td>
<td>✓</td>
</tr>
<tr>
<td>S10</td>
<td>79</td>
<td>4</td>
<td>250</td>
<td>8</td>
<td>R</td>
<td>HA</td>
<td>75</td>
<td>✓</td>
</tr>
<tr>
<td>S11</td>
<td>80</td>
<td>3</td>
<td>900</td>
<td>8</td>
<td>R</td>
<td>HA</td>
<td>75</td>
<td>✓</td>
</tr>
<tr>
<td>S12</td>
<td>61</td>
<td>7</td>
<td>900</td>
<td>8</td>
<td>R</td>
<td>-</td>
<td>75</td>
<td>✓</td>
</tr>
<tr>
<td>Mean (N=12)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S13</td>
<td>70</td>
<td>5</td>
<td>1200</td>
<td>8</td>
<td>L</td>
<td>HA</td>
<td>75</td>
<td>✓</td>
</tr>
<tr>
<td>S14</td>
<td>72</td>
<td>3</td>
<td>1200</td>
<td>8</td>
<td>L</td>
<td>HA</td>
<td>75</td>
<td>✓</td>
</tr>
<tr>
<td>S4</td>
<td>46</td>
<td>6</td>
<td>1800</td>
<td>8</td>
<td>L</td>
<td>HA</td>
<td>75</td>
<td>✓</td>
</tr>
<tr>
<td>S15</td>
<td>75</td>
<td>6</td>
<td>900</td>
<td>8</td>
<td>L</td>
<td>-</td>
<td>75</td>
<td>✓</td>
</tr>
<tr>
<td>S16</td>
<td>58</td>
<td>8</td>
<td>250</td>
<td>8</td>
<td>R</td>
<td>-</td>
<td>75</td>
<td>✓</td>
</tr>
<tr>
<td>S17</td>
<td>67</td>
<td>4</td>
<td>900</td>
<td>8</td>
<td>L</td>
<td>-</td>
<td>75</td>
<td>✓</td>
</tr>
<tr>
<td>S18</td>
<td>74</td>
<td>2</td>
<td>1200</td>
<td>7</td>
<td>R</td>
<td>HA</td>
<td>75</td>
<td>✓</td>
</tr>
<tr>
<td>S1</td>
<td>54</td>
<td>3</td>
<td>900</td>
<td>8</td>
<td>R</td>
<td>HA</td>
<td>75</td>
<td>✓</td>
</tr>
<tr>
<td>S7</td>
<td>67</td>
<td>4</td>
<td>900</td>
<td>8</td>
<td>L</td>
<td>HA</td>
<td>75</td>
<td>✓</td>
</tr>
<tr>
<td>S19</td>
<td>79</td>
<td>11</td>
<td>900</td>
<td>8</td>
<td>R</td>
<td>-</td>
<td>75</td>
<td>✓</td>
</tr>
<tr>
<td>S20</td>
<td>69</td>
<td>3</td>
<td>900</td>
<td>10</td>
<td>L</td>
<td>-</td>
<td>75</td>
<td>✓</td>
</tr>
<tr>
<td>Mean (N=11)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S21</td>
<td>77</td>
<td>3</td>
<td>900</td>
<td>6</td>
<td>R</td>
<td>-</td>
<td>65</td>
<td>✓</td>
</tr>
<tr>
<td>S14</td>
<td>73</td>
<td>4</td>
<td>1200</td>
<td>8</td>
<td>L</td>
<td>HA</td>
<td>65</td>
<td>✓</td>
</tr>
<tr>
<td>S22</td>
<td>62</td>
<td>4</td>
<td>900</td>
<td>8</td>
<td>R</td>
<td>HA</td>
<td>65</td>
<td>✓</td>
</tr>
<tr>
<td>S23</td>
<td>62</td>
<td>3</td>
<td>900</td>
<td>8</td>
<td>L</td>
<td>HA</td>
<td>65</td>
<td>✓</td>
</tr>
<tr>
<td>S5</td>
<td>44</td>
<td>5</td>
<td>900</td>
<td>8</td>
<td>L</td>
<td>-</td>
<td>65</td>
<td>✓</td>
</tr>
<tr>
<td>S8</td>
<td>70</td>
<td>5</td>
<td>900</td>
<td>8</td>
<td>R</td>
<td>HA</td>
<td>65</td>
<td>✓</td>
</tr>
<tr>
<td>S20</td>
<td>70</td>
<td>4</td>
<td>900</td>
<td>10</td>
<td>L</td>
<td>-</td>
<td>65</td>
<td>✓</td>
</tr>
<tr>
<td>S24</td>
<td>74</td>
<td>6</td>
<td>900</td>
<td>8</td>
<td>L</td>
<td>HA</td>
<td>65</td>
<td>✓</td>
</tr>
<tr>
<td>Mean (N=8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Experiment I: Mean 65.6 3.3

Experiment II: Mean 66.5 5.0

Experiment III: Mean 66.5 4.2
### 3.9 Experiment summary by outcome measure and noise type

<table>
<thead>
<tr>
<th>Subject</th>
<th>Age (yrs)</th>
<th>Implant use (yrs)</th>
<th>Stimulation rate (Hz)</th>
<th>Number of maxima</th>
<th>Test ear</th>
<th>Non-test ear device</th>
<th>C-SPL</th>
<th>ADRO</th>
</tr>
</thead>
<tbody>
<tr>
<td>S25</td>
<td>74</td>
<td>7</td>
<td>720</td>
<td>12</td>
<td>R</td>
<td>HA</td>
<td>65</td>
<td>✓</td>
</tr>
<tr>
<td>S24</td>
<td>74</td>
<td>6</td>
<td>900</td>
<td>8</td>
<td>L</td>
<td>HA</td>
<td>65</td>
<td>✓</td>
</tr>
<tr>
<td>S26</td>
<td>64</td>
<td>3</td>
<td>900</td>
<td>8</td>
<td>R</td>
<td>HA</td>
<td>65</td>
<td>✓</td>
</tr>
<tr>
<td>S7</td>
<td>67</td>
<td>5</td>
<td>900</td>
<td>8</td>
<td>L</td>
<td>HA</td>
<td>65</td>
<td>✓</td>
</tr>
<tr>
<td>S15</td>
<td>76</td>
<td>7</td>
<td>900</td>
<td>8</td>
<td>L</td>
<td>-</td>
<td>65</td>
<td>✓</td>
</tr>
<tr>
<td>S27</td>
<td>79</td>
<td>12</td>
<td>250</td>
<td>8</td>
<td>L</td>
<td>-</td>
<td>65</td>
<td>✓</td>
</tr>
<tr>
<td>S10</td>
<td>81</td>
<td>6</td>
<td>250</td>
<td>8</td>
<td>R</td>
<td>HA</td>
<td>65</td>
<td>✓</td>
</tr>
<tr>
<td>S28</td>
<td>74</td>
<td>1</td>
<td>900</td>
<td>8</td>
<td>R</td>
<td>HA</td>
<td>65</td>
<td>✓</td>
</tr>
<tr>
<td>S1</td>
<td>56</td>
<td>6</td>
<td>900</td>
<td>8</td>
<td>R</td>
<td>HA</td>
<td>65</td>
<td>✓</td>
</tr>
<tr>
<td>S21</td>
<td>77</td>
<td>3</td>
<td>900</td>
<td>6</td>
<td>R</td>
<td>-</td>
<td>65</td>
<td>✓</td>
</tr>
<tr>
<td>S29</td>
<td>54</td>
<td>3</td>
<td>900</td>
<td>8</td>
<td>R</td>
<td>HA</td>
<td>65</td>
<td>✓</td>
</tr>
<tr>
<td>S30</td>
<td>50</td>
<td>4</td>
<td>900</td>
<td>8</td>
<td>R</td>
<td>HA</td>
<td>65</td>
<td>✓</td>
</tr>
<tr>
<td>S31</td>
<td>80</td>
<td>6</td>
<td>500</td>
<td>6</td>
<td>R</td>
<td>-</td>
<td>65</td>
<td>✓</td>
</tr>
</tbody>
</table>

**Mean (N=13)** 70.0 5.3

<table>
<thead>
<tr>
<th>Subject</th>
<th>Age (yrs)</th>
<th>Implant use (yrs)</th>
<th>Stimulation rate (Hz)</th>
<th>Number of maxima</th>
<th>Test ear</th>
<th>Non-test ear device</th>
<th>C-SPL</th>
<th>ADRO</th>
</tr>
</thead>
<tbody>
<tr>
<td>S25</td>
<td>74</td>
<td>7</td>
<td>720</td>
<td>12</td>
<td>R</td>
<td>HA</td>
<td>65</td>
<td>✓</td>
</tr>
<tr>
<td>S32</td>
<td>72</td>
<td>4</td>
<td>900</td>
<td>8</td>
<td>R</td>
<td>HA</td>
<td>65</td>
<td>✓</td>
</tr>
<tr>
<td>S21</td>
<td>77</td>
<td>3</td>
<td>900</td>
<td>6</td>
<td>R</td>
<td>-</td>
<td>65</td>
<td>✓</td>
</tr>
<tr>
<td>S4</td>
<td>46</td>
<td>6</td>
<td>1800</td>
<td>8</td>
<td>L</td>
<td>HA</td>
<td>75</td>
<td>✓</td>
</tr>
<tr>
<td>S14</td>
<td>73</td>
<td>4</td>
<td>1200</td>
<td>8</td>
<td>L</td>
<td>HA</td>
<td>75</td>
<td>✓</td>
</tr>
<tr>
<td>S33</td>
<td>57</td>
<td>1</td>
<td>900</td>
<td>8</td>
<td>R</td>
<td>HA</td>
<td>65</td>
<td>✓</td>
</tr>
<tr>
<td>S22</td>
<td>62</td>
<td>4</td>
<td>900</td>
<td>8</td>
<td>R</td>
<td>HA</td>
<td>65</td>
<td>✓</td>
</tr>
<tr>
<td>S28</td>
<td>74</td>
<td>1</td>
<td>900</td>
<td>8</td>
<td>R</td>
<td>HA</td>
<td>65</td>
<td>✓</td>
</tr>
<tr>
<td>S29</td>
<td>54</td>
<td>3</td>
<td>900</td>
<td>8</td>
<td>R</td>
<td>HA</td>
<td>65</td>
<td>✓</td>
</tr>
<tr>
<td>S17</td>
<td>67</td>
<td>4</td>
<td>900</td>
<td>8</td>
<td>L</td>
<td>-</td>
<td>75</td>
<td>✓</td>
</tr>
<tr>
<td>S23</td>
<td>62</td>
<td>3</td>
<td>900</td>
<td>8</td>
<td>L</td>
<td>HA</td>
<td>65</td>
<td>✓</td>
</tr>
<tr>
<td>S30</td>
<td>50</td>
<td>4</td>
<td>900</td>
<td>8</td>
<td>R</td>
<td>HA</td>
<td>65</td>
<td>✓</td>
</tr>
<tr>
<td>S18</td>
<td>74</td>
<td>2</td>
<td>1200</td>
<td>7</td>
<td>R</td>
<td>HA</td>
<td>75</td>
<td>✓</td>
</tr>
<tr>
<td>S24</td>
<td>74</td>
<td>6</td>
<td>900</td>
<td>8</td>
<td>L</td>
<td>HA</td>
<td>65</td>
<td>✓</td>
</tr>
<tr>
<td>S2</td>
<td>73</td>
<td>2</td>
<td>900</td>
<td>12</td>
<td>L</td>
<td>HA</td>
<td>65</td>
<td>✓</td>
</tr>
</tbody>
</table>

**Mean (N=15)** 65.9 3.6
Chapter 4

Experiment I

Bias value in low reverberation

The main objectives of this experiment were to identify the intelligibility benefit due to the SpatialNR algorithm and to study the effect of the bias parameter used to adjust the algorithm’s aggressiveness. Performance was compared against the Beam baseline in a low reverberation sound treated room using SRT tests in one noise type (4-TB-rear).

4.1 Introduction

Experiment I was designed as the first clinical investigation of the SpatialNR algorithm with CI recipients. The main objective of Experiment I was to assess the benefit, if any, of SpatialNR in terms of speech intelligibility in noisy conditions. The gain threshold parameter was expected to vary the aggressiveness of noise reduction by changing the SNR-dependant attenuation of noisy channels. The gain threshold parameter was tested across a range of settings to understand the impact on speech intelligibility.

4.2 Method

The following section describes the methods used in this experiment. The specific details of parameter settings for the processing conditions are described. An assessment of the theoretical performance of the processing conditions is included. The specific acoustic test conditions, outcome measures and test protocol used in the experiment are given.
Table 4-1: Configuration of SpatialNR processing conditions evaluated in Experiment I

<table>
<thead>
<tr>
<th>Label</th>
<th>Input signal directionality</th>
<th>Bias $\alpha$ (dB)</th>
<th>Smoothing $\beta_\alpha/\beta_\beta$ (ms)</th>
<th>Max Attenuation $\gamma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SpatialNR(S, 0)</td>
<td>Standard</td>
<td>0</td>
<td>10/10</td>
<td>0 (-inf dB)</td>
</tr>
<tr>
<td>SpatialNR(Z, -3)</td>
<td>Zoom</td>
<td>-3</td>
<td>10/10</td>
<td>0 (-inf dB)</td>
</tr>
<tr>
<td>SpatialNR(Z, 0)</td>
<td>Zoom</td>
<td>0</td>
<td>10/10</td>
<td>0 (-inf dB)</td>
</tr>
<tr>
<td>SpatialNR(Z, +3)</td>
<td>Zoom</td>
<td>3</td>
<td>10/10</td>
<td>0 (-inf dB)</td>
</tr>
<tr>
<td>SpatialNR(Z, +6)</td>
<td>Zoom</td>
<td>6</td>
<td>10/10</td>
<td>0 (-inf dB)</td>
</tr>
</tbody>
</table>

4.2.1 Processing conditions

The evaluation was performed with a group of 12 CI listeners by comparing SpatialNR processing against the Beam baseline.

Several variations of the SpatialNR algorithm were evaluated in this study. In particular, bias values, $\alpha$, were evaluated at 2, 1, 0.5, and 0.25 corresponding to biasing $\xi dB$ by -3 dB, 0 dB, +3 dB, and +6 dB SNR respectively. In addition to varying the bias, a second algorithmic variable was tested where the noise reduction gains were applied to the Standard directional signal as opposed to the Zoom signal. This variation was tested using $\alpha = 1$ (0 dB) to form a total of five SpatialNR conditions. The details are shown in Table 4-1 and are abbreviated as follows: SpatialNR (S, 0), SpatialNR (Z, -3), SpatialNR (Z, 0), SpatialNR (Z, +3), and SpatialNR (Z, +6), where S represents Standard, Z represents Zoom, and the bias value is in units of dB.

Theoretical performance of SpatialNR under different bias parameter settings and input signal directionality are shown in Figure 4-1. There is an inherent trade-off in type I and type II errors as the bias value is increased – the type I errors decrease while the type II errors increase. The optimal trade-off for CI users is difficult to predict, although the STI indicates that bias value of 0-3dB is likely to be optimal. There appears to be a clear benefit of using the Zoom directionality as opposed to Standard, judging by the lower type I and type II errors, and higher STI achieved with Zoom as the input signal directionality.

The Beam condition was evaluated to gauge benchmark performance. Beam, an adaptive beamformer (Spriet et al., 2007) commercially available in the CP810 speech processor, was used as a baseline for comparison because it has demonstrated superior noise reduction performance to other directional microphone options available in the device (Hersbach et al., 2012).

4.2.2 Acoustic conditions

4-TB-rear noise was selected as the noise type for Experiment I and evaluation was performed in the Whitford sound booth at Cochlear, Melbourne. This selection was based on the assumption that both Beam and SpatialNR should provide benefit by attenuating noise from behind the listener, and was therefore likely to be a scenario in which both algorithms should provide an advantage over the Standard directionality.
4.2 Method

Theoretical performance in 4-TB-rear
SpatialNR bias value

![Graph showing theoretical performance](image)

**Figure 4-1:** Theoretical performance of SpatialNR algorithm with different bias value and input signal directionality, measured in 4-TB-rear noise in the sound treated Whitford booth. Increasing the bias value provides more aggressive noise reduction.
4.2.3 Outcome measures

Speech intelligibility was evaluated using SRT.

4.2.4 Test protocol

Evaluation took place over two separate visits to the laboratory for each of the 12 recipients. The processing conditions were tested in random order in the first session then in reverse order in the second session providing counterbalancing. SRTs were averaged over the two sessions.

4.3 Results

SRTs for each subject are shown in Figure 4-2 for the six processing conditions evaluated, and the SRT benefit relative to the Beam baseline is shown in Figure 4-3 for each of the five SpatialNR processing conditions summarized for the group of 12 subjects. The benefit relative to Beam was computed by subtracting the SpatialNR SRT from the Beam SRT for each individual, and is shown in Figure 4-3 as box and whisker plots. The mean SRT for the Beam condition was 0.3 dB.
4.3 Results

Figure 4-3: SRT benefit relative to Beam in spatially separated 4-talker babble from the rear hemi-field. Box plots show the 10, 25, 50, 75 and 90\textsuperscript{th} percentiles and all outliers are marked with dots. The means are shown as dotted lines. Stars indicate statistically significance ($P<0.001$).

Analysis of the group results using a two-way repeated-measures analysis of variance (RM-ANOVA, Shavelson, 1988) with processing condition and test session as factors revealed the following:

- a significant main effect of processing condition ($F[5,55]=41.69$, $P<0.001$),
- no significant main effect of test session ($P=0.324$) and
- no significant interaction term (processing condition x test session, $P=0.984$).

This indicates that training or acclimatization was unlikely to have affected the SRTs recorded at the two test sessions.

Post-hoc pairwise analyses using Student-Newman-Keuls methods showed that all SpatialNR conditions were significantly better than Beam ($P<0.001$) regardless of the bias parameter value or whether the SpatialNR filter weights were applied to the Standard or Zoom signal. SpatialNR (S, 0) showed 2.7 dB (SE=0.4) benefit over Beam ($P<0.001$). Changing the signal to which the filter weights were applied from Standard to Zoom provided a 1.5 dB (SE=0.4, $P<0.001$) improvement such that SpatialNR (Z, 0) was 4.2 dB (SE=0.2) better than Beam.

---

\textsuperscript{9} In ANOVA, $F[n,m]$ is the test statistic with n and m degrees of freedom drawn from the F-distribution (Shavelson, 1988, Ch. 14). $P$ is the corresponding level of significance for rejecting the null hypothesis based on the comparison of the F-statistic to a table of critical F-values. A $P$-value less than or equal to 0.05 is considered significant for rejecting the null hypothesis at the 5\% significance level, thus concluding a systematic effect that could not be observed due to random sampling alone.
4.4.1 SpatialNR parameter settings

Two variable aspects to the SpatialNR algorithm were investigated: changing the input signal directionality to which the filter weights were applied (Standard/Zoom) and changing the bias parameter, \( \alpha \). For all variants, the performance was significantly better than Beam, and SpatialNR (Z, +3) demonstrated the best improvement of 4.6 dB SRT. This agrees with the

\( P<0.001 \). The effect of changing the bias parameter from -3 dB to +3 dB resulted in a performance benefit of 1.2 dB (SE=0.5, \( P<0.05 \) whilst all other comparisons of bias parameter settings failed to reach a significant difference. A bias value of +3 dB (\( \alpha=0.5 \)) showed the largest improvement over the Beam condition of 4.6 dB (SE=0.4, \( P<0.001 \)).

Results at the individual level show there was a range of different performers within the group, indicated by the range of SRTs recorded for the baseline Beam condition. The range covered more than 15 dB, from -4.4 dB to +10.5 dB. A Pearson product moment correlation was performed to analyse the relationship between the baseline Beam SRT and the improvement due to each of the SpatialNR conditions, however no statistically significant relationship was found (\( P=0.065 \) SpatialNR(S, 0), \( P=0.137 \) SpatialNR(Z, -3), \( P=0.067 \) SpatialNR(Z, 0), \( P=0.452 \) SpatialNR(Z, +3), \( P=0.895 \) SpatialNR (Z, +6)).

In order to analyse the relationship between bias value and test SNR, a Pearson product moment correlation was performed between the Beam SRT and the bias value which resulted in the largest improvement for each subject; however, no significant correlation was found (\( P=0.189 \)).

### 4.4 Discussion

The SpatialNR (Z, +3) algorithm showed a statistically significant improvement in SRTs of 4.6 dB over the Beam algorithm. Given that Beam has previously been shown to provide a 5.0 dB improvement in SRTs compared to the Standard directionality setting in a different group of subjects (Hersbach et al. 2011), the overall improvement of SpatialNR over the Standard setting is likely to have increased speech intelligibility from 0% to 100% in the noise environment used for evaluation, and so could provide substantial improvements for CI users in commonly encountered noisy situations.

The main reason for superior performance of SpatialNR over Beam is the ability to remove multiple interfering noise sources at different locations simultaneously. Beam must adapt a filter to minimize the total energy in the output and while it is possible to obtain very large attenuation of a single masker, the effectiveness is reduced when multiple noise sources are presented simultaneously from different locations (van Hoesel and Clark, 1995; Hamacher et al., 1997; Wouters and Vanden Berghe, 2001). In contrast, the SpatialNR algorithm can cancel many sources simultaneously from many locations because all sources contribute to a common SNR estimate, which is used to remove the noise.
theoretical measurement of STI, which indicates the highest STI was achieved at a bias value of +3 dB.

**SpatialNR input signal**

The choice of signal to which the SpatialNR gains were applied significantly affected speech intelligibility, with the choice of Zoom being 1.5 dB better than Standard. Theoretical measurements of STI and error rates also support this. This benefit is due to the spatial pre-processing of the Zoom signal, which has a directivity pattern that attenuates signals from the rear more heavily than the Standard configuration. The difference is smaller than measured in a direct comparison between Standard and Zoom directionality without the SpatialNR post-filter, which resulted in a 3.9 dB difference in SRT in the same noise environment\(^{10}\) (Hersbach et al., 2012). This indicates that there is some overlap in the noise removed by Zoom directionality and SpatialNR, and indicates some redundancy in the information used by the two algorithms.

**SpatialNR bias parameter**

Adjusting the bias value provided a means of exchanging the relative amount of noise reduction and speech distortion in the output signal, therefore altering the aggressiveness of the system. As the bias value (in dB) was increased, the amount of noise reduction was increased (as indicated by type I error reduction), but more speech was also removed as a consequence leading to speech distortion (as indicated by type II error increase). The bias parameter value had some impact on performance. In addition, not all parameter values that were evaluated resulted in statistically significant performance differences compared with one another, the maximum benefit over the Beam condition was found to be 4.6 dB with bias value of +3 dB (α = 0.5). The bias value greater than zero indicates CI users tolerated some speech distortion in the signal, suggesting that trading type I errors for type II errors was preferential. A bias value of -3 dB (α = 2) was significantly worse than a bias value of +3 dB, an expected outcome given the degree of noise attenuation was reduced, allowing more noise to pass through the system. Although there was no significant difference between other values of the bias parameter, increasing the bias value beyond +6 dB would likely to lead to decreased performance due to excessive speech distortion. This hypothesis is supported by theoretical measurements, although it cannot be verified with the present SRT data. However, bias parameter values beyond +6 dB could be useful in other noise environments where there are interfering maskers closer to the target signal. In such cases, the bias parameter could be used to increase noise reduction while maintaining acceptable speech distortion and, therefore, a different value of bias may be optimal. This was explored further in Experiment II (Chapter 5).

\(^{10}\) The noise environment was identical apart from the timing of the spatial randomisation changes. In the current study, the changes were controlled to occur three seconds before each sentence began and did not change during the sentence. In Hersbach et al. (2012) the changes occurred at regular three second intervals, not synchronised to the start of each sentence, and therefore mid-sentence changes to the spatial configuration were possible.
4.4.2 Noise type and test SNR

The large improvements demonstrated are expected to have significant benefit for CI users since the noise environment under which the evaluation took place can be considered a reasonably difficult condition for understanding speech, and may simulate some difficult real-world environments. During adaptive SRT testing, the SpatialNR algorithm was evaluated across a wide range of SNRs spanning more than 15 dB due to the different CI users within the group. No significant relationship between the baseline Beam SRT and SpatialNR SRTs was found, which indicates the test SNR did not have a significant impact on the performance benefit obtained with SpatialNR. In addition, no significant relationship was found between the baseline Beam SRT and the bias value which provided maximum benefit for each subject, indicating that the best bias parameter was not dependent on the test SNR. It appears the algorithm is robust to changes in the input SNR supporting its suitability for a wide range of listening environments. However, the locations of interfering talkers were restricted to the rear hemi-field and, since both Beam and SpatialNR algorithms are based on spatial filtering, it is expected that the benefit would be reduced when maskers are located closer to the target and would be least effective when the target and masker are co-located. In this instance, single-channel noise reduction is a viable alternative, and has been shown to provide speech intelligibility benefit, at least in situations where the masker is less modulated than the target signal (Dawson et al., 2011; Hersbach et al., 2012).

Although substantial noise reduction performance was demonstrated, such aggressive and successful noise reduction may have some detrimental outcomes in daily situations. This may become apparent when the target signal is not directly in front of the listener, or when sound from behind is not considered entirely as interfering noise. Common situations occur when a CI user is interested in signals both from in front and behind. SpatialNR is unlikely to be suitable for this kind of scenario and, in order to become effective in a commercial device, manual or automatic switching of the algorithm is likely to be required to accommodate the wishes of the user. This was explored further in Experiment V (Chapter 8).

4.4.3 Reverberation

The sound-treated room used for presentation of the noisy environment had low reverberation, and it is well known that adaptive beamformer performance declines as reverberation increases. This is directly linked to the spatial response of the noise reference, which ideally has a deep null in the target direction but under the effects of reverberation the null depth is reduced. Further evaluation of the SpatialNR algorithm in the presence of reverberation will assist in assessing the likely effect on day-to-day use of the algorithm in real-world environments. This was explored further in Experiment II (Chapter 5).

4.4.4 Acclimatisation

No effect of training or acclimatisation was measured across the two test sessions for any of the processing conditions evaluated in this study. This indicates that training was unlikely to have had a significant influence on SRTs gathered in this study. However, listening experience
is known to play a key role affecting CI performance when new sound processing strategies are evaluated, and further exposure to the test processing conditions could lead to increased benefit for individual listeners. This was explored further in Experiment V (Chapter 8).

4.5 Conclusion

A multiple-microphone noise reduction algorithm (SpatialNR) was evaluated in a complex noisy environment where the noise sources changed location during the test. Speech intelligibility tests with CI recipients revealed a significant benefit of SpatialNR processing compared to a commercially available adaptive beamformer. This benefit was likely due to the design of SpatialNR, which allowed simultaneous removal of multiple noise sources that is not possible with the adaptive beamformer. The bias parameter of the SpatialNR algorithm provided a mechanism for trading off speech distortion and noise reduction, and was used to control the aggressiveness of noise reduction. The benefit obtained with the experimental algorithm was large, comparable to the benefit observed when a second microphone was introduced into Nucleus CI systems.
Chapter 5

Experiment II

Bias value and time constants in reverberation

In this experiment, the noise type used for evaluation introduced more competing talkers (20-TB-full) and a moderate level of reverberation. Bias value and time constant parameters were assessed by measuring both speech intelligibility and sound quality in two types of noise (4-TB-rear and 20-TB-full). Testing in quiet was used to highlight any negative aspects of processing associated with removal of the target signal under such conditions. The baseline processing conditions were expanded to include Standard, Zoom and Beam.

5.1 Introduction

In Experiment I, an initial evaluation of the SpatialNR algorithm was performed in order to establish a first demonstration of benefit that could be provided by the algorithm. Evaluation took place in a sound treated room with low reverberation in one noise condition, 4-TB-rear. The results demonstrated that SpatialNR could produce substantial benefit over currently available directional microphone technology. Experiment II was designed to further consolidate the benefits by introducing reverberation and more diffuse noise types, whilst further exploring a range of parameter settings to understand their impact on performance over a range of acoustic environments. It was anticipated that as the level of reverberation was increased, the benefit of SpatialNR would be reduced due to the uncorrelated signals received at the microphones. Introduction of noise sources into the front hemi-field was also anticipated to reduce benefit, given the reduced spatial separation of target and noise.
5.2 Algorithm parameter selection

The algorithm parameters that were varied in this experiment were the bias parameter and smoothing time constants. Theoretical performance metrics were used to refine the range of parameter values that were used in the perceptual evaluation with CI users.

The bias parameter was varied based on results from Experiment I (Chapter 4) and on theoretical performance measures, with the expectation that there might be dependence of the optimal choice of bias value based on the noise configuration. The range of parameters chosen was +3, +6 and +9 dB and the theoretical performance is shown in Figure 5-1. The theoretical performance data suggest that increasing the bias value will lead to fewer type I errors (improved noise reduction) at the expense of increasing type II errors (increasing speech distortion), and therefore there is a trade-off between the two. The STI figures indicate best performance at bias values of +3 and +6 dB in 4-TB-rear and 20-TB-full noise respectively, and degraded performance as the bias value is increased to +9 dB in both noise types due to increased speech distortion (type II errors). Similarly, decreasing the bias value also leads to decreased STI, due to less removal of background noise.

Further to the bias parameter modifications, asymmetrical smoothing of the signal and noise estimates was introduced with 5/50 ms attack/release times. The alternative smoothing was partially based on theoretical performance regarding error rates (Figure 5-2) although differences between the settings that were clinically evaluated produced relatively little impact on error rates, indicated with open and closed markers in Figure 5-2. Informal normal hearing listening tests provided further input to the selection, which suggested that the most satisfactory sound quality occurred with asymmetric smoothing using 5/50 ms. Therefore, both error rates and subjective listening tests were taken into consideration when selecting the smoothing time constants used in the perceptual evaluation with CI listeners.

5.3 Method

The following section describes the methods used in this experiment. Details of specific parameter settings for each processing condition are provided. The acoustic test conditions, outcome measures and test protocol are given. Table 5-2 summarises the processing conditions and outcome measures used in the experiment.
Figure 5-1: Theoretical performance of SpatialNR (smoothing 5/50 ms) with different bias value. Calculated in 4-TB-rear and 20-TB-full noise at 0 dB SNR in the reverberant training room.
Theoretical performance for SpatialNR smoothing parameters

4-TB-rear

20-TB-full

Type I

Release time [ms]

0 2 5 7.5 20 50 75 200

10/10 ms

5/50 ms

Type II

Attack time [ms]

0 2 5 7.5 10 20 50 75 200

Figure 5-2: Theoretical performance of SpatialNR for different smoothing parameters (using fixed bias value +3 dB) in 4-TB-rear (left panels) and 20-TB-full (right panels) noise types. Heat maps show type I errors (upper panels) and type II errors (lower panels) with higher error rates shaded darker and contour intervals of 1%. For reference, open and closed points indicate the 10/10 and 5/50 ms combinations used in clinical testing. Acoustic stimulus was presented in the NAL Training room using running speech in noise at 0 dB SNR.
Table 5-1: Configuration of SpatialNR processing conditions evaluated in Experiment II

<table>
<thead>
<tr>
<th>Label</th>
<th>Input signal directionality</th>
<th>Bias $\alpha$ (dB)</th>
<th>Smoothing $\beta_a/\beta_r$ (ms)</th>
<th>Max Attenuation $\gamma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SpatialNRa</td>
<td>Zoom</td>
<td>3</td>
<td>10/10</td>
<td>0 (-inf dB)</td>
</tr>
<tr>
<td>SpatialNRb</td>
<td>Zoom</td>
<td>3</td>
<td>5/50</td>
<td>0 (-inf dB)</td>
</tr>
<tr>
<td>SpatialNRc</td>
<td>Zoom</td>
<td>6</td>
<td>5/50</td>
<td>0 (-inf dB)</td>
</tr>
<tr>
<td>SpatialNRd</td>
<td>Zoom</td>
<td>9</td>
<td>5/50</td>
<td>0 (-inf dB)</td>
</tr>
</tbody>
</table>

Table 5-2: Summary of processing conditions and outcome measures used in the evaluation during Experiment II.

<table>
<thead>
<tr>
<th>Processing condition label</th>
<th>CP810 directionality setting</th>
<th>Experimental algorithm</th>
<th>Experimental algorithm parameters</th>
<th>AuSTIN</th>
<th>MUSHRA</th>
<th>CNC’s</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard (no reverb)</td>
<td>Standard</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Standard</td>
<td>Standard</td>
<td>-</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Zoom</td>
<td>Zoom</td>
<td>-</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Beam</td>
<td>Beam</td>
<td>NR</td>
<td>Gain threshold $=+3$ dB</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>SpatialNRa</td>
<td>Zoom</td>
<td>SpatialNR</td>
<td>Bias $=+3$ dB Smoothing $=10/10ms$</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SpatialNRb</td>
<td>Zoom</td>
<td>SpatialNR</td>
<td>Bias $=+3$ dB Smoothing $=5/50ms$</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>SpatialNRc</td>
<td>Zoom</td>
<td>SpatialNR</td>
<td>Bias $=+6$ dB Smoothing $=5/50ms$</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>SpatialNRd</td>
<td>Zoom</td>
<td>SpatialNR</td>
<td>Bias $=+9$ dB Smoothing $=5/50ms$</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5.3.1 Processing conditions

Three directionality configurations, Standard, Zoom and Beam, were tested as baseline conditions. The CP900 SNR-NR algorithm was evaluated in combination with the Beam directionality setting (BeamNR), whilst SpatialNR was evaluated in combination with Zoom directionality as the input signal. The signal smoothing time constant and bias parameter settings of SpatialNR were varied to understand the impact on speech intelligibility and sound quality in the reverberant environment. A bias value of $+3$ dB was used to evaluate two different smoothing configurations which differed in the attack and release time constants used. SpatialNRa used 10/10 ms (attack/release) and SpatialNRb used 5/50 ms (attack/release), both with bias value of $+3$ dB. Three bias parameter values ($+3$, $+6$, and $+9$ dB) were evaluated with the 5/50 ms (attack/release) smoothing configuration (SpatialNRb,
SpatialNRc, and SpatialNRd respectively). Details of parameter configuration for the SpatialNR conditions are shown in Table 5-1.

5.3.2 Acoustic conditions

Two noise types were selected for evaluation. 4-TB-rear was chosen to maintain a comparison with Experiment I and 20-TB-full as a more difficult environment with more competing talkers. The stimuli were presented in the reverberant training room with the intention that it would be representative of a real room rather than the artificially damped environment of the sound treated booth. The target-to-listener distance was 1 m and the noise-to-listener distance was 3 m. All processing was performed using the impulse response data measured in the training room. An additional control condition was included for the Whitford sound booth with low reverberation\(^\text{11}\).

5.3.3 Outcome measures

Three outcome measures were collected as part of the evaluation. They were:

- SRT speech intelligibility in noise,
- MUSHRA subjective sound quality in noise and
- CNC word recognition in quiet.

5.3.4 Test protocol

The study used a repeated-measure, single-subject design, in which each subject served as his/her own control. In total, nine conditions were available for testing; Standard (no reverb), Standard, Zoom, Beam, BeamNR, SpatialINRa, SpatialINRb, SpatialNRc, SpatialNRd. All nine conditions were evaluated via the AuSTIN SRT test over the course of two test sessions for each noise type, so there was four test sessions in total. The test order of processing conditions within the first session for each noise type was randomised and then reversed for the second session of the same noise type. The order of testing of different noise types was also randomised and counterbalanced within each subject across the four test sessions such that the session order was ABBA or BAAB, where A and B were the different noise types. For the MUSHRA sound quality and CNC word tasks, the number of test conditions was reduced by selecting only variants of SpatialINR that were expected to produce the best speech intelligibility in the AuSTIN SRT task and therefore SpatialINRa and SpatialINRd were not retained. The Standard (no reverb) condition was also removed as a test condition for the MUSHRA sound quality task. The MUSHRA task was completed at two SNRs (-5, +5 dB SNR) in each test session and averaged across the two test sessions with the same noise type\(^\text{12}\).

\(^\text{11}\) The target and noise distances for the Whitford booth were not the same as those in the training room, but were set to 1.2 m due to physical dimensions of the room.

\(^\text{12}\) Although not reported herein, bilateral noise reduction algorithms were evaluated during the same clinical trial in order to make most efficient use of clinical resources and volunteer subjects’ time. The SRT evaluation had four bilateral noise reduction conditions, the MUSHRA sound quality task and CNC word test had two additional
5.4 Results

The results are presented in sections which relate to the different outcome measures evaluated in the experiment.

5.4.1 SRT

The group mean SRTs are shown in Figure 5-3 for the group of CI users under each of the processing conditions. The two-way RM-ANOVA with noise type and processing condition as factors revealed a significant interaction between noise type and processing condition ($P<0.001$) indicating that algorithm performance depended on the noise type that was used for evaluation. Student-Newman-Keuls procedure for multiple pairwise comparisons was used in the following post-hoc analysis.

Comparison of SpatialNR conditions to Beam as a baseline was performed (Figure 5-4) as a relevant demonstration of the potential benefit of SpatialNR over the best performing commercially available processing. In 4-TB-rear, all SpatialNR conditions were significantly better than Beam (all $P<0.001$) and the greatest improvement was 4.3 dB due to SpatialNRb. In 20-TB-full, the benefit due to SpatialNR was smaller than 4-TB-rear, and SpatialNRc provided the greatest improvement over Beam by 1.6 dB ($P<0.001$) while SpatialNRb provided a marginally non-significant 1 dB benefit over Beam ($P=0.064$). The two SpatialNR parameters that were altered in this evaluation were smoothing time-constant (SpatialNRa vs. SpatialNRb) and bias parameter (SpatialNRb vs. SpatialNRc vs. SpatialNRd). However, over the range of parameter values evaluated, changes failed to produce a statistically significant difference in SRTs in both 4-TB-rear and 20-TB-full noise types.

Comparison of the baseline directionality settings (Standard, Zoom and Beam) in 4-TB-rear revealed a significant improvement due to Zoom (over Standard) of 4.1 dB ($P<0.001$) and a significant improvement due to Beam (over Zoom) of a further 1.3 dB ($P<0.001$). In 20-TB-full, there was a significant improvement of 2.8 dB ($P<0.001$) due to Zoom (over Standard) and no further benefit due to Beam (over Zoom, $P=0.720$).

In order to assess the benefit of the single-channel NR algorithm, a comparison of the Beam and BeamNR conditions was made. The comparison revealed no significant effect due to NR processing in 4-TB-rear ($P=0.359$), nor in 20-TB-full ($P=0.079$).

A rather unexpected result was obtained in the reverberant room compared to the sound treated room. Comparison of the conditions Standard (no reverb) and Standard revealed the group of CI recipients performed better in the reverberant room. The difference was 1.1 dB ($P=0.003$) in 4-TB-rear, and 1.5 dB ($P<0.001$) in 20-TB-full.

---

5.4 Results

The results are presented in sections which relate to the different outcome measures evaluated in the experiment.

5.4.1 SRT

The group mean SRTs are shown in Figure 5-3 for the group of CI users under each of the processing conditions. The two-way RM-ANOVA with noise type and processing condition as factors revealed a significant interaction between noise type and processing condition ($P<0.001$) indicating that algorithm performance depended on the noise type that was used for evaluation. Student-Newman-Keuls procedure for multiple pairwise comparisons was used in the following post-hoc analysis.

Comparison of SpatialNR conditions to Beam as a baseline was performed (Figure 5-4) as a relevant demonstration of the potential benefit of SpatialNR over the best performing commercially available processing. In 4-TB-rear, all SpatialNR conditions were significantly better than Beam (all $P<0.001$) and the greatest improvement was 4.3 dB due to SpatialNRb. In 20-TB-full, the benefit due to SpatialNR was smaller than 4-TB-rear, and SpatialNRc provided the greatest improvement over Beam by 1.6 dB ($P<0.001$) while SpatialNRb provided a marginally non-significant 1 dB benefit over Beam ($P=0.064$). The two SpatialNR parameters that were altered in this evaluation were smoothing time-constant (SpatialNRa vs. SpatialNRb) and bias parameter (SpatialNRb vs. SpatialNRc vs. SpatialNRd). However, over the range of parameter values evaluated, changes failed to produce a statistically significant difference in SRTs in both 4-TB-rear and 20-TB-full noise types.

Comparison of the baseline directionality settings (Standard, Zoom and Beam) in 4-TB-rear revealed a significant improvement due to Zoom (over Standard) of 4.1 dB ($P<0.001$) and a significant improvement due to Beam (over Zoom) of a further 1.3 dB ($P<0.001$). In 20-TB-full, there was a significant improvement of 2.8 dB ($P<0.001$) due to Zoom (over Standard) and no further benefit due to Beam (over Zoom, $P=0.720$).

In order to assess the benefit of the single-channel NR algorithm, a comparison of the Beam and BeamNR conditions was made. The comparison revealed no significant effect due to NR processing in 4-TB-rear ($P=0.359$), nor in 20-TB-full ($P=0.079$).

A rather unexpected result was obtained in the reverberant room compared to the sound treated room. Comparison of the conditions Standard (no reverb) and Standard revealed the group of CI recipients performed better in the reverberant room. The difference was 1.1 dB ($P=0.003$) in 4-TB-rear, and 1.5 dB ($P<0.001$) in 20-TB-full.

---

conditions. Therefore, for the purpose of statistical analysis and to avoid detection of false positives as real differences, p-values are reported for RM-ANOVAs using the entire data set. This means that the reported p-values are higher and represent extra caution used in reporting significant differences between conditions.

13 This difference of 1 dB was significant in the reduced RM-ANOVA in which the bilateral conditions were removed from the analysis ($P=0.037$).

14 In the RM-ANOVA with reduced set of processing conditions the effect of NR was still non-significant, the P values were 0.360 and 0.071 for 4-TB-rear and 20-TB-full noise types respectively.
Figure 5-3: Group mean SRTs in 4-talker rear half (4-TB-rear) and 20-talker full circle (20-TB-full) noise for all processing conditions evaluated in Experiment II. Lower scores indicate better performance. Error bars show the 95% confidence interval.
5.4 Results

5.4.2 MUSHRA

The group mean rating scores from the MUSHRA task are shown in Figure 5-5. After excluding the reference condition from the analysis, a three-way RM-ANOVA using noise, processing and SNR as factors was performed. All main effects were found to be significant (noise $P=0.002$, processing $P<0.001$ and SNR $P<0.001$) and all interactions were not significant (noise x processing $P=0.650$, noise x SNR $P=0.142$, processing x SNR $P=0.462$, and noise x processing x SNR $P=0.597$). Therefore, for the purposes of analysis, each main effect was analysed separately after averaging the data over the other two factors.

Comparisons for the processing factor revealed both SpatialNR conditions (SpatialNRb and SpatialNRc) were rated significantly better than all other conditions ($P<0.001$ for all comparisons), and not significantly different from each other ($P=0.895$). SpatialNRc was rated most highly (apart from the Reference condition) with a mean rating of 61, followed by SpatialNRb (60), Zoom (46) and Standard (36). The difference of 10 rating points between Zoom and Standard was significant ($P=0.005$).

---

The results for conditions involving the adaptive Beam algorithm (Beam and BeamNR) are not presented due to lack of time to adapt within the MUSHRA test framework. Although subjects were able to listen to up to 12 seconds of stimulus, subjects tended to make their rating decision faster than this, often within the first few seconds. With an adaptation time of 3-5 seconds, full adaptation may not have occurred before sound quality ratings were made.
Figure 5-5: Group mean sound quality ratings for the MUSHRA task in 4-talker rear half (4-TB-rear) and 20-talker full circle (20-TB-full) noise at two different input SNRs. Error bars indicate the 95% confidence interval. SpatialNRa and SpatialNRd were omitted from evaluation in the MUSHRA task. The Reference condition used Standard processing at an input SNR of +20 dB. Data from Beam and BeamNR conditions were not able to be interpreted due to lack of adaptation of the Beam algorithm within the test framework.
5.5 Discussion

5.5.1 SpatialNR

The SpatialNR algorithm improved speech understanding and sound quality for the group of CI users listening to speech embedded in spatially separated babble noise presented in a reverberant room. Intelligibility performance compared to Beam was improved by 4.1 dB SRT in 4-TB-rear configuration. In Experiment I, which took place with a different group of CI listeners in a low reverberation sound treated booth, the benefit of SpatialNR was 4.6 dB SRT compared to Beam when evaluated in 4-TB-rear. This demonstrates that the benefit of the SpatialNR algorithm was maintained in the reverberant environment used for evaluation.

Figure 5-6: Group mean CNC word scores in quiet. Error bars indicate 95% confidence intervals. SpatialNRa and SpatialNRd were omitted from CNC word testing.

For the noise factor, the mean rating was 51 in 4-TB-rear, and 42 in 20-TB-full, a significant difference of 9 rating points between noise types. For the SNR factor, the mean rating was 30 at -5 dB SNR and 63 at +5 dB SNR, a significant difference of 33 rating points.

5.4.3 CNCs

The group mean CNC word scores in quiet are shown in Figure 5-6. A one-way RM-ANOVA revealed no significant effect of processing condition (P=0.349).

5.5 Discussion
The benefit of SpatialNR in 20-TB-full noise was 1.6 dB SRT in comparison to Beam, indicating SpatialNR was still effective at attenuating competing noise with less spatial separation from the target speech. The benefit was reduced compared to 4-TB-rear, an expected finding due to the location of competing talkers closer to the target speech. Sound quality in 4-TB-rear was rated 9 points higher than 20-TB-full and there was no significant interaction between noise type and processing condition. This indicates the quality improvement due to SpatialNR was not dependent on the noise type used for evaluation, which contrasts with the speech intelligibility findings, where benefit in 4-TB-rear was superior to the benefit in 20-TB-full. Sound quality ratings at the two test SNRs (-5 and +5 dB SNR) were rated differently by 30 rating points, and there was no significant interaction between SNR and processing condition. This indicates that quality improvement due to SpatialNR was not dependent on the test SNR. The significant improvement in speech intelligibility and sound quality demonstrated in both noise environments that included reverberation, as well as robustness over the SNRs used in the evaluation, suggests SpatialNR is likely to provide benefits for CI users in real environments.

Changing the bias and smoothing parameters of the SpatialNR algorithm failed to produce statistically significant changes in SRTs or sound quality ratings, although maximum intelligibility was achieved with different bias parameter settings in the two different noise types. This trend was also detected in the theoretical STI, which was maximum at 3 dB bias in 4-TB-rear, and maximum at 6 dB bias in 20-TB-full. In general, different bias settings may be optimal for different noise configurations, but evaluation confirmed that a parameter choice of +3 dB bias and 5/50 ms smoothing is suitable for noise reduction across these two noise configurations.

5.5.2 Zoom/Beam

In the current study, the benefit relative to Standard directionality in the reverberant room with 4-TB-rear was 4.1 dB and 5.6 dB, for Zoom and Beam respectively. The present study allowed for a direct comparison of SRTs using Standard directionality with and without reverberation, but not a direct comparison of Zoom and Beam in different reverberation. However, in Experiment I, which used a different group of subjects and a low reverberation sound booth, the reported SRT benefit was 3.8 dB for Zoom and 5.0 dB for Beam in comparison to Standard directionality. Comparison with the present findings indicates that the directional benefit was maintained in the reverberant room used for evaluation in this study, and that the room acoustics did not have a significant detrimental effect on the ability of these algorithms to reduce noise. It is generally accepted that reverberation has a negative effect on the performance of adaptive beamformers (Greenberg and Zurek, 2001), however the target to listener distance and noise to listener distance need to be considered in the context of the reverberant space. The reverberation time of the room alone may not be sufficient to predict performance.
5.5.3 NR

Comparison of SRTs for Beam and BeamNR showed that NR did not produce a significant benefit in either noise type, indicating that modulated noise could not be estimated accurately enough by the single channel technique. This is a common finding, as previously demonstrated where performance improved in speech weighted steady noise, but not in 4-talker or 20-talker babble (Hersbach et al., 2012). It was hypothesised that the introduction of reverberation may lead to better noise estimation due to temporal smearing of the noise envelope, but this did not eventuate in a statistically significant improvement in SRTs in this experiment, although the comparison was only marginally non-significant at the 95% level (P=0.071).

5.5.4 CI performance in reverberation

A direct comparison of performance in the Standard condition in two rooms with different reverberation showed that CNC word scores in quiet were not significantly affected by reverberation, but SRTs were significantly improved. That is, in the presence of background noise, increased reverberation led to an improvement in performance. This is possibly due to the fact that in the reverberant room, the target speech presented from 1 m was largely unaffected by the reflective surfaces in the room. This was demonstrated through equivalent word recognition in quiet conditions in the two rooms. The short early decay time (EDT) at the 1 m position and high values of clarity and direct-to-reverberant ratio (C50, C7 and DRR) confirm the low reverberant energy compared to the direct sound for the target speech signal. The noise presented from 3 m was more affected by the room acoustics compared to the target speech, reflected in the longer EDT, and lower values of C50, C7 and DRR. The performance increase could be attributed to the fact that the target speech was largely unaffected by the room acoustics while the competing babble was temporally smeared due to reverberation. Therefore, the competing babble was less modulated in the reverberant condition than in the low reverberation condition. Unlike their NH hearing counterparts, CI listeners show an advantage when competing noise is less modulated (Nelson et al., 2003), which may explain why performance improved with increased reverberation in this study.

5.5.5 Future research

In this study, all input signals were generated via convolution with impulse responses. Therefore, it was not possible to assess the perceptual effect of head movement during the operation of SpatialNR. In addition, processing of the user’s own voice was not assessed during the evaluation. These are important aspects that will impact the clinical success or otherwise of the algorithm.

5.6 Conclusion

SpatialNR was evaluated in moderately reverberant conditions, and provided significant benefit over Beam and other directional processing algorithms. In fact, the introduction of
reverberation did not appear to reduce the benefit that SpatialNR, Beam or Zoom could provide. The comparison of Standard directionality across rooms with different reverberant properties suggests that CI users may benefit from increased reverberation, at least for the configuration where the target is reasonably close to the listener. A range of parameter variation was evaluated for bias and smoothing parameters, which led to the choice of +3 dB and 5/50 ms respectively as optimal. These parameters are suitable for further evaluations, and are likely to lead to maximum benefit across a wide range of acoustic conditions.
Chapter 6

Experiment III

High reverberation

In this experiment, the effects of reverberation were explored by altering the distance of target and noise from the listener in a reverberant room. High levels of reverberation were tested using a room simulator. A baseline omni-directional processing condition was introduced while algorithm parameters remained fixed for this experiment. Evaluation was performed using SRT tests in one noise type (4-TB-full).

6.1 Introduction

In Experiment II, evaluation of sound processing algorithms was performed in reverberant conditions to establish if there was any impact of reverberation on CI performance generally, and to establish if there was any impact of reverberation on the benefit that multi-microphone algorithms could provide. The experiment revealed that the group of CI recipients performed better as reverberation was increased and that multi-microphone processing was largely unaffected by increasing the reverberant energy in the competing noise signal. Experiment III was designed to investigate these aspects further, by making a more direct comparison between reverberant and non-reverberant conditions, and to further increase the level of reverberation to a point where the benefit of multi-microphone noise reduction was reduced. This was achieved through a room simulator used to create an extreme reverberation condition. The reverberant room simulator was validated as part of the experiment.

6.2 Methods

In the following sections, the methods used in the experiment are presented. The room simulator operation and configuration is described and compared to the reverberant
properties of the real room. The sound processing conditions are given and specific SpatialNR parameter settings are provided and summarised in Table 6-2. The acoustic test conditions describe the reverberation level and distances between the target/noise and the listener, summarised in Table 6-3. The outcome measures and test protocol are also provided.

6.2.1 Room simulator

The room simulator used was MCRoomSim (Wabnitz et al., 2010). It was a shoebox simulator that modelled the reflection and scattering of sound waves from six surfaces (floor, ceiling, 4 x walls). The simulator was based on the work of Schimmel et al. (2009) with an important extension that made the simulator suitable for use in these experiments. The extension could simulate microphone arrays with arbitrary directional impulse responses, and provided realistic phase information for the sound recorded by the microphone array, including accurate inter-sensor time delays (Wabnitz et al., 2010). This allowed the use of the simulator in these experiments, in which the impulse responses were used to generate multiple microphone signals that were then subject to further processing, thus requiring accurate simulation of the inter-microphone response.

The simulator was configured by defining the absorption and scattering coefficients for each of the six surfaces of the simulated shoebox room. The sound processor microphone polar responses were modelled using the anechoic response of the microphones when worn on a KEMAR manikin. In this way, the effects of the head and microphone location behind the ear were modelled. The location of the stimulus source and the nature of the acoustic device producing the stimulus were also modelled.

The reverberation simulator was used to produce a set of impulse responses that modelled the training room by setting the coefficients of absorption and scattering based on published material properties of the floor, ceiling and wall surfaces. The dimensions of the room, the location of KEMAR within the room and the location of the loudspeakers were also modelled to replicate the training room impulse response recordings. In addition, a highly reverberant condition “Simulated(Concrete)” was simulated by changing the properties of the acoustic surfaces in the model to simulate concrete walls, ceiling and floors, perhaps similar to an undercover car park environment. The simulator was used to produce a set of impulse responses at each listening distance and position.

With the choice of reverberation properties, the energy decay curve was calculated from the recorded and simulated impulse responses shown in Figure 6-1. The reverberant properties were extracted from the impulse responses and calculated at all three distances (1m, 2m, 3m) and 12 angles (from 0 to 330 degrees in steps of 30 degrees) that were simulated. Table 6-1 shows the acoustic properties of the Real, Simulated and Simulated(Concrete) after averaging over the 12 angles at each listening distance.

The Schroeder energy decay curve (Schroeder, 1979) and reverberation properties were well matched between the real and simulated impulse responses, given the selection of acoustic properties based on published data.
Figure 6-1: The energy decay curves comparing the measured and simulated impulse responses from the front left microphone at a distance of (A) 1, (B) 2 and (C) 3 m.
Table 6-1: Acoustic properties of the training room, calculated from the measured and simulated impulse responses. The data presented were calculated as the median over each set of 12 values (calculated at angles from 0 to 330 degrees, in 30 degree increments)

<table>
<thead>
<tr>
<th>Distance (m)</th>
<th>Impulse Response</th>
<th>T30 (ms)</th>
<th>EDT (ms)</th>
<th>C50 (dB)</th>
<th>C7 (dB)</th>
<th>DRR (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Real</td>
<td>460</td>
<td>20</td>
<td>17.2</td>
<td>12.0</td>
<td>11.1</td>
</tr>
<tr>
<td></td>
<td>Simulated</td>
<td>622</td>
<td>15</td>
<td>18.4</td>
<td>14.9</td>
<td>14.7</td>
</tr>
<tr>
<td></td>
<td>Simulated (Concrete)</td>
<td>809</td>
<td>432</td>
<td>11.6</td>
<td>9.8</td>
<td>9.8</td>
</tr>
<tr>
<td>2</td>
<td>Real</td>
<td>528</td>
<td>402</td>
<td>11.8</td>
<td>5.9</td>
<td>4.5</td>
</tr>
<tr>
<td></td>
<td>Simulated</td>
<td>696</td>
<td>284</td>
<td>13.0</td>
<td>9.4</td>
<td>6.8</td>
</tr>
<tr>
<td></td>
<td>Simulated (Concrete)</td>
<td>861</td>
<td>966</td>
<td>6.7</td>
<td>4.2</td>
<td>3.0</td>
</tr>
<tr>
<td>3</td>
<td>Real</td>
<td>581</td>
<td>412</td>
<td>9.2</td>
<td>3.3</td>
<td>-0.3</td>
</tr>
<tr>
<td></td>
<td>Simulated</td>
<td>742</td>
<td>589</td>
<td>10.3</td>
<td>6.4</td>
<td>3.8</td>
</tr>
<tr>
<td></td>
<td>Simulated (Concrete)</td>
<td>887</td>
<td>1007</td>
<td>4.1</td>
<td>0.9</td>
<td>-0.4</td>
</tr>
</tbody>
</table>

Table 6-2: Configuration of SpatialNR processing condition evaluated in Experiment III

<table>
<thead>
<tr>
<th>Label</th>
<th>Input signal directionality</th>
<th>Bias $\alpha$ (dB)</th>
<th>Smoothing $\beta_s/\beta_a$ (ms)</th>
<th>Max Attenuation $\gamma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SpatialNR</td>
<td>Zoom</td>
<td>3</td>
<td>5/50</td>
<td>0 (-inf dB)</td>
</tr>
</tbody>
</table>

6.2.2 Processing conditions

Omni, Zoom, Beam and SpatialNR were all evaluated. SpatialNR had a gain threshold of +3 dB and smoothing time constant of 5/50 ms attack/release, chosen as the parameters that provided consistent benefit over the range of noise environments evaluated in Experiment II. The detailed parameter configuration for SpatialNR used in Experiment III is shown in Table 6-2.

6.2.3 Acoustic conditions

Using the recorded impulse responses from the training room, the reverberant properties of the speech and noise components of the signal were altered by changing the distance from the listening position. By moving a source further away from the listening position, the DRR was decreased, since there was more reverberant energy in the signal reaching the microphone. Three configurations were created, which changed the reverberant properties of the speech and noise, by locating the speech and noise at distances of either 1 m or 3 m. At 1 m distance, the DRR was 11.1 dB, and at 3 m, the DRR was -0.3 m, just outside the critical distance at which the direct to reverberant energy was equal. The three configurations were named Low R, Mid R and High R, indicating the level of reverberant energy in the signal (Table 6-3).
Table 6-3: Physical separation of target and noise source from the listener and method of impulse response generation for each of the five reverberant configurations used for evaluation in Experiment III

<table>
<thead>
<tr>
<th>Reverberation Level</th>
<th>Target Distance [m]</th>
<th>Noise Distance [m]</th>
<th>Impulse Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low R</td>
<td>1</td>
<td>1</td>
<td>Real</td>
</tr>
<tr>
<td>Mid R</td>
<td>1</td>
<td>3</td>
<td>Real</td>
</tr>
<tr>
<td>High R</td>
<td>3</td>
<td>3</td>
<td>Real</td>
</tr>
<tr>
<td>HighSim</td>
<td>3</td>
<td>3</td>
<td>Simulated</td>
</tr>
<tr>
<td>ExtraHighSim</td>
<td>3</td>
<td>3</td>
<td>Simulated(Concrete)</td>
</tr>
</tbody>
</table>

In order to validate the MCRoomSim simulator, impulse responses were generated using the simulator modelling the High R condition. The simulator modelled the acoustic properties of the training room with the sources for the speech and noise located at a distance of 3 m and this condition was called HighSim (Table 6-3).

Finally, a condition with very high reverberation was created with the simulator by modelling the acoustic surfaces as concrete, and this condition was called ExtraHighSim (Table 6-3).

6.2.4 Outcome measures

SRTs were evaluated in 4-TB-full noise type.

6.2.5 Test protocol

There were 20 test conditions in total, comprised of four sound processing conditions and five reverberation configurations. Given the large number of test conditions, only one SRT per condition was collected per condition for each subject. The order of testing was randomised for each subject, and spread across two visits to the laboratory.

Protocol deviations

S24: during the first visit, the test condition Zoom in High R was mistakenly tested with SpatialNR processing. The subject was not able to be retested; therefore, this data point was missing from the analysis.

6.3 Results

The results are presented in sections which address the effect of physical sound source distance from the listener, validation of the room simulator and influence of very high reverberation.
Figure 6-2: SRT results for each processing condition in (A) Low R, (B) Mid R and (C) High R reverberation conditions. Error bars show the 95% C.I.

Figure 6-3: SRT Benefit over the Omni processing conditions. Data has been averaged across three reverberation conditions Low R, Mid R and High R. Error bars show the 95% C.I.
6.3 Results

6.3.1 Physical distance to sound sources

The comparison between Low R, Mid R and High R was made to analyse the effect of the physical distance between the target/noise sources and the listener within the reverberant room. The mean SRTs for the Low R, Mid R and High R conditions are shown in Figure 6-2. The two-way RM-ANOVA with reverberation level and processing as factors revealed:

- a significant main effect of reverberation level ($F[2,14]=23.23, P<0.001$),
- a significant main effect of processing condition ($F[3,21]=80.80, P<0.001$) and
- no significant interaction between the main factors ($F[6,42]=1.43, P=0.227$).

This indicates that the effect of processing was not dependent on the level of reverberation. To further test this, the benefit over omni was calculated, and the two-way RM-ANOVA was re-run. The analysis revealed:

- a significant main effect of processing ($P=0.004$),
- no significant main effect of reverberation ($P=0.098$), and
- the interaction term was not significant ($P=0.659$).

This shows that while overall CI performance varied with reverberation, the benefit of directional processing was relatively consistent over the Low, Mid and High levels of reverberation.

Post-hoc comparisons (Student-Newman-Keuls) of the reverberation factor showed a 0.8 dB difference between the Low R and Mid R conditions which was not significant ($P=0.095$). However, the High R condition (with target speech at 3 m) resulted in significantly worse performance of 2.1 dB ($P<0.001$) compared to Low R and 2.8 dB ($P<0.001$) compared to Mid R. Comparison of the processing condition factor showed that all pairs except Beam vs. Zoom resulted in significantly different comparisons. In particular, the benefit over omni (Figure 6-3) was significant for Zoom (4.2 dB, $P<0.001$), Beam (3.7 dB, $P<0.001$) and SpatialNR (5.0 dB, $P<0.001$). The benefit of SpatialNR over Zoom was 0.8 dB ($P=0.043$) and over Beam was 1.3 dB ($P=0.005$).

6.3.2 Room simulator validation

The comparison between High R and HighSim was made to validate the room simulator. The group SRTs for these are shown in Figure 6-4 for each processing condition. The two-way RM-ANOVA with reverberation level and processing condition as main factors revealed no significant effect of reverberation level ($F[1,7]=4.30, P=0.076$). As expected, a significant main effect of processing condition ($F[3,21]=43.23, P<0.001$) was found. The interaction term, reverberation x processing was not significant ($F[3,21]=0.791, P=0.513$). Post-hoc comparisons revealed the level of reverberation did not produce a significant difference in performance in any of the processing conditions (omni $P=0.937$, Zoom $P=0.112$, Beam $P=0.111$ and SpatialNR $P=0.937$).
Figure 6-4: SRT Results for (A) High R and (B) HighSim reverberation conditions. Error bars indicate 95% C.I.

Figure 6-5: SRT results for HighSim and ExtraHighSim reverberation conditions. Error bars indication the 95% C.I. Note the change of scale from previous figures with similar data.
6.3 Results

Figure 6-6: SRT benefit of each processing condition over the Omni conditions, in HighSim and ExtraHighSim levels of reverberation. Error bars indicate 95% C.I.

6.3.3 Very high reverberation

The comparison between reverberation levels HighSim and ExtraHighSim was made to analyse the effect of room acoustics on speech intelligibility. The group SRTs are shown in Figure 6-5. The two-way RM-ANOVA with reverberation level and processing as main factors revealed a significant interaction term, reverberation x processing (F[3,21]=5.68, P=0.006). This indicates that the effect of changing the reverberation level had a different impact on the different types of processing. For all processing conditions except SpatialNR, the ExtraHighSim condition significantly reduced speech intelligibility performance (i.e. resulted in higher SRTs). The ExtraHighSim condition reduced performance by 9.6 dB in the Omni condition (P<0.001), 7.7 dB in the Zoom condition (P<0.001), 5.3 dB in the Beam condition (P=0.002) and was not significant in the SpatialNR condition (P=0.102). Generally, the ExtraHighSim produced results that were more variable across the subject group than the HighSim condition.

In order to understand the impact of reverberation on the benefit that each algorithm provided, the SRT benefit over Omni was calculated at each level of reverberation (Figure 6-6). Consistent with the results for High R presented in section 0, each processing condition provided a significant benefit over Omni in the HighSim condition. The benefit was 5.0 dB for Zoom (P=0.016), 4.5 dB for Beam (P=0.013) and 4.1 dB for SpatialNR (P=0.006). There was no significant difference between any of the conditions Zoom, Beam and SpatialNR in HighSim. The benefit of each algorithm over Omni was maintained in ExtraHighSim and the benefit was significant for all conditions. The benefit over Omni in ExtraHighSim was 6.4 dB for Zoom (P<0.001), 8.8 dB for Beam (P<0.001) and 11.2 dB for SpatialNR (P<0.001). In order to test if the SRT benefit in ExtraHighSim was significantly different from the benefit in HighSim, a two-way RM-ANOVA on the SRT benefit over omni using processing condition and reverberation level as factors was performed. It revealed a significant interaction between the SRT benefit
over Omni and the level of reverberation, consistent with the analysis of raw SRTs. Post-hoc analysis showed the SRT benefit over Omni in each of the processing conditions increased with ExtraHighSim compared to HighSim. The SRT benefit over Omni due to SpatialNR increased significantly by 7.1 dB (P=0.005). The increase in SRT benefit over omni due to Zoom (2.1 dB, P=0.369) and Beam (4.3 dB, P=0.053) failed to reach significance due to the high variability in results across the group in the ExtraHighSim condition.

6.4 Discussion

The discussion is presented similarly to the results, separated by an analysis of the physical sound source distance from the listener, validation of the room simulator and influence of very high reverberation.

6.4.1 Physical distance to sound sources

Within the NAL training room, changing the level of reverberant energy in the signal at the microphones was initially achieved by altering the source to listener distance within the room. Comparing Low R and Mid R conditions indicates the effect of changing the distance of the noise sources from 1 m to 3 m whilst keeping the target distance constant at 1 m. The comparison revealed a small (0.8 dB) non-significant impact on intelligibility (P=0.095). A similar pattern of results was evident in Experiment II, which showed a significant improvement in reverberation of 1.1 dB in 4-TB-rear and 1.5 dB in 20-TB-full noise types. A possible mechanism underlying this small effect could lie in the amplitude modulations of the competing babble that were reduced as the reverberant energy increased. The higher levels of reverberant energy filled in the fluctuations in amplitude of the competing babble, making it less modulated overall. CI recipients were able to make use of the temporally smeared noise signal, avoiding the negative effects associated with modulated maskers likely to be associated with the nature of envelope based processing strategies (such as ACE) and/or the lack of spectral detail in the presented stimuli (Nelson et al., 2003). This finding suggest that moderate levels of reverberation are potentially beneficial to speech understanding of CI users, so long as the target speech remains close to the listener, and the reverberant energy is mostly influenced by competing noise.

In the High R condition, with speech and noise both at a distance of 3 m, the mean performance of the group of CI subjects was significantly reduced compared to the Low R and Mid R conditions. That is, moving the target speech from 1 m to 3 m resulted in decreased performance. This suggests that the reverberant energy of the target speech played a significant role in speech intelligibility in the reverberant room when background babble noise was present, and perhaps had more impact than the competing babble noise itself. The effect was consistent across Omni, Zoom, Beam and SpatialNR such that the pattern of results obtained with the different algorithms was not significantly affected by changing the distance between sound sources and the listening position within the reverberant room. SpatialNR provided the most benefit in comparison with the Omni condition, and was also significantly better than Beam and Zoom. These results suggest that the target to listener distance may play
a key role in speech understanding in reverberant conditions where background noise is present.

6.4.2 Room simulator validation

Validation of the room simulator was performed by comparing the High R and HighSim results. There was no significant effect of changing the real impulse responses for the simulated impulse responses, and the performance of each algorithm was maintained with the change. This demonstrates that the simulator was useful for producing two-microphone impulse responses with relevant inter-microphone transfer functions that led to realistic processing of the microphone signals thereafter. The simulated room acoustic properties produced impulse responses of approximately equal reverberant properties to the real room. This finding suggests that the room simulator could be used to generate impulse responses for arbitrary rooms, with some confidence that the process generates meaningful impulse response data. Indeed it allowed us to analyse the ExtraHighSim data with some level of confidence that the simulator was equivalent to recordings made in a real room of equivalent acoustic properties.

6.4.3 Very high reverberation

Comparing HighSim and ExtraHighSim conditions showed the effect of introducing concrete surfaces into the room. The comparison showed that performance in the Omni condition was significantly reduced with the introduction of concrete surfaces. This finding is supported by other studies which show that CI performance in noise is made more difficult as the level of reverberation is increased (Poissant et al., 2006; Hazrati and Loizou, 2012). The comparison also showed that the performance in the Zoom and Beam conditions were significantly reduced with the introduction of concrete surfaces whilst performance of SpatialNR was not significantly different. However, the performance benefit due to each algorithm compared with Omni was maintained for Zoom and Beam, and increased for SpatialNR in the concrete room. This demonstrates that all of the directional algorithms provided useful benefit in the highly reverberant condition at least on par with the benefit obtained in lower reverberation. SpatialNR provided the most benefit, and the degree of benefit significantly increased as the reverberation was increased with the introduction of concrete surfaces into room. This finding is in contrast to other works, which contend that high levels of reverberation are detrimental to the performance of multi-microphone noise reduction algorithms (e.g. van Hoesel and Clark, 1995; Hamacher et al., 1997; Greenberg and Zurek, 2001). The current finding is not completely without precedence, however. Leeuw and Dreschler (1991) found that while hearing aid performance decreased with listening distance in a reverberant room (RT60=900ms), the benefit due to directionality was not significantly affected as the listening distance was increased. Ricketts and Hornsby (2003) studied the effects of reverberation and listening distance on hearing aid directional microphones and found that, for a given listening distance, the level of reverberation (RT60 300ms and 900ms) did not affect directional microphone benefit, and only in the room with higher reverberation did listening distance significantly affect performance. Therefore, it is likely that the spatial configuration of the competing noise and physical location of target and noise sources within the reverberant room...
play an important role in the outcome and ultimate benefit that multi-microphone processing is able to provide.

The results of this experiment suggest that not only the acoustic properties of the room, but also the physical location of target and noise sources within the room, particularly the target to listener distance, play an important role in speech understanding for CI listeners. The benefit provided by all directional algorithms was maintained even with reasonable high levels of reverberation, suggesting that directional noise reduction algorithms, and in particular SpatialNR, provide robust benefits in terms of speech understanding in noisy reverberant environments.

6.5 Conclusion

Testing was performed in a range of reverberant environments. There was evidence that the physical distance of target and noise sound sources within the reverberant room play an important role in the ability of CI users to understand speech. In particular, a small target distance and larger noise distance appear to be beneficial due to the relative level of reverberant energy introduced into the two signals. A software room simulator was used to create a room with concrete walls, in which CI performance was severely degraded. The benefit of two-microphone noise reduction algorithms, Zoom, Beam and SpatialNR, were tested and compared against the Omni condition. While CI performance in the Omni condition degraded at larger target distances, the benefit provided by the noise reduction algorithms was not significantly affected by changing the target and noise source distance within the limits of the test conditions. The concrete room produced surprising results where the benefit of Zoom and Beam was at least on par with the benefit in lower reverberation, and the benefit of SpatialNR was significantly increased. The findings suggest that directional microphone algorithms were beneficial in noisy reverberant conditions, that the benefit was increased as reverberation was increased and that SpatialNR provided the most benefit of all algorithms that were evaluated.
Chapter 7

Experiment IV

Microphone mismatch

Microphone mismatch is known to distort the spatial response of directional microphones. In this experiment, algorithm’s which rely on the directional microphone patterns were investigated by increasing the degree of mismatch between the microphones. Speech intelligibility and sound quality were measured in one noise type (4-TB-full).

7.1 Introduction

In Experiments I-III, the SpatialNR algorithm was evaluated in range of acoustic environments that varied in the level of reverberation that was present. The multi-microphone noise reduction algorithms that were evaluated (Standard, Zoom, Beam and SpatialNR) generally provided robust benefits that were not detrimentally affected by increased level of reverberation. Although the algorithms were robust, the level of reverberation represents a departure from ideal acoustic conditions which are worthy of investigation to understand performance in the real world. Another source of non-ideal conditions is associated with the properties of the microphones themselves, in particular the sensitivity and phase matching across the frequency range used for processing. The design of multi-microphone noise reduction algorithms often works upon the assumption that the two (or more) microphones are ideally matched. However, in practice, there are limitations to how close this matching is, and microphone properties can change over time. Therefore, for a practical device, it is important to understand the impact of model errors associated with microphone mismatch. Experiment IV was designed to address this issue by introducing a controlled broad band sensitivity mismatch and measuring the resulting speech understanding and sound quality.
7.2 Methods

The following sections describe the methods used in this experiment. In particular the level of microphone mismatch used is described with an accompanying analysis of the theoretical performance under those mismatch conditions. The processing conditions, acoustic conditions, outcome measures and test protocol are also provided.

7.2.1 Microphone mismatch

Microphone mismatch was introduced in a controlled fashion by attenuating the rear microphone within the real-time sound processing model. Gain was applied to the time domain samples, and thus produced a broad-band sensitivity mismatch between the two microphones. The level of attenuation was set to 0, 2, 4 and 6 dB during the clinical trial, with the intention of determining the relationship between broad band sensitivity mismatch and performance.

7.2.2 Theoretical performance

The introduced mismatch was expected to affect the performance of all algorithms to some extent, due to the influence of the mismatch on the fixed directional patterns that were utilised by all algorithms. The influence on Zoom and noise reference spatial patterns is shown in Figure 7-1, which indicates the directional patterns become more omni-directional as the mismatch is increased.

Theoretical performance of each processing condition under different levels of microphone mismatch was evaluated using error rate and STI metrics, shown in Figure 7-2. The data were calculated after presenting running speech in 4-TB-full noise at 0 dB SNR in the Whitford sound treated booth. The data indicate that as microphone mismatch is increased, both type I and type II error rates are increased and STI is decreased. This suggests that clinical performance is likely to be reduced as microphone mismatch is increased.

7.2.3 Processing conditions

Standard, Zoom, Beam and SpatialNR were all evaluated. SpatialNR had a gain threshold of +3 dB and smoothing time constant of 5/50 ms attack/release, chosen as the parameters which provided consistent benefit over the range of noise environments evaluated in Experiment II. The detailed parameter configuration for SpatialNR used in Experiment IV is shown in Table 7-1.
### Table 7-1: Configuration of SpatialNR processing condition evaluated in Experiment IV

<table>
<thead>
<tr>
<th>Label</th>
<th>Input signal directionality</th>
<th>Bias $\alpha$ (dB)</th>
<th>Smoothing $\beta_A/\beta_R$ (ms)</th>
<th>Max Attenuation $\gamma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SpatialNR</td>
<td>Zoom</td>
<td>3</td>
<td>5/50</td>
<td>0 (-inf dB)</td>
</tr>
</tbody>
</table>

---

**Figure 7-1:** Broad-band spatial response of (A) Zoom and (B) noise reference with different levels of microphone mismatch indicated with shading. The plots show the directional sensitivity in dB under anechoic conditions when the microphones are worn on KEMAR manikin left ear. For each level of microphone mismatch, the sensitivity at each angle is relative to 0 degrees (Zoom) or 180 degrees (noise reference).
Theoretical performance in 4-TB-full

Type I error rate

Type II error rate

STI

Figure 7-2: (A) Type I error rate, (B) type II error rate and (C) speech transmission index (STI) for various levels of microphone mismatch. Noise type used was 4-TB-full at 0 dB SNR in the Whitford sound treated booth.
7.2 Methods

7.2.4 Acoustic conditions

Evaluation took place in 4-TB-full noise type in the sound treated room.

7.2.5 Outcome measures

Outcome metrics used were SRT and sound quality (MUSHRA).

7.2.6 Test protocol

For SRT testing, microphone mismatch was set to 0, 2 and 4 dB, whilst in MUSHRA testing an additional mismatch level (6 dB) was added to explore rather large mismatch, but was not included in SRT testing due to time limitations. Therefore, for SRT testing, there were 12 test conditions in total, comprising three levels of microphone mismatch and four different processing conditions. SRTs were collected over two test sessions for each subject. The test order was randomised in the first session, and reversed in the second session to provide counter-balancing of the test order across the two sessions. SRTs were averaged over the two test sessions.

MUSHRA testing involved splitting the test conditions into separate evaluations since 12 conditions (plus reference and hidden reference) was too many to compare simultaneously. Therefore, separate comparisons were made for Standard, Zoom, Beam and SpatialNR, with an additional baseline comparison as outlined in Table 7-2. Each comparison included an Anchor (Standard processing at 0 dB mismatch), Visible Reference, Hidden Reference and processing at mismatch levels of 0, 2, 4 and 6 dB using the chosen processing condition. An additional baseline comparison was made in order to link the different processing conditions. In that comparison, all of the processing conditions were compared at 0 dB mismatch. The baseline condition was necessary to make comparisons of absolute ratings across different conditions, since absolute ratings reported in the MUSHRA test can be affected by the other conditions to which they are compared (Zielinski et al., 2007). For example, the absolute rating of a particular condition can be different if the other conditions are all of a relatively high quality compared to situations where the other conditions are of a relatively poor quality. Therefore, five MUSHRA comparisons were made in the first test session (one for each processing condition and one baseline) and these were repeated in the second session. The sound quality ratings were averaged across the two sessions. For subjects unfamiliar with the MUSHRA task, a practice run was administered prior to data collection.

Protocol deviations

There were two protocol deviations during SRT testing. One of two SRT runs was tested with the incorrect mismatch settings for S26 (SpatialNR-0dB) and S30 (Standard-2dB). Therefore, only a single SRT contributed to the average for these two data points.
Table 7-2: Processing conditions that were compared simultaneously for each instance of the MUSHRA sound quality task

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Condition Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>Reference</td>
</tr>
<tr>
<td></td>
<td>Standard-0dB</td>
</tr>
<tr>
<td></td>
<td>Zoom-0dB</td>
</tr>
<tr>
<td></td>
<td>Beam-0dB</td>
</tr>
<tr>
<td></td>
<td>SpatialNRb-0dB</td>
</tr>
<tr>
<td>Standard</td>
<td>Reference</td>
</tr>
<tr>
<td></td>
<td>Anchor</td>
</tr>
<tr>
<td></td>
<td>Standard-0dB</td>
</tr>
<tr>
<td></td>
<td>Standard-2dB</td>
</tr>
<tr>
<td></td>
<td>Standard-4dB</td>
</tr>
<tr>
<td></td>
<td>Standard-6dB</td>
</tr>
<tr>
<td>Zoom</td>
<td>Reference</td>
</tr>
<tr>
<td></td>
<td>Anchor</td>
</tr>
<tr>
<td></td>
<td>Zoom-0dB</td>
</tr>
<tr>
<td></td>
<td>Zoom-2dB</td>
</tr>
<tr>
<td></td>
<td>Zoom-4dB</td>
</tr>
<tr>
<td></td>
<td>Zoom-6dB</td>
</tr>
<tr>
<td>Beam</td>
<td>Reference</td>
</tr>
<tr>
<td></td>
<td>Anchor</td>
</tr>
<tr>
<td></td>
<td>Beam-0dB</td>
</tr>
<tr>
<td></td>
<td>Beam-2dB</td>
</tr>
<tr>
<td></td>
<td>Beam-4dB</td>
</tr>
<tr>
<td></td>
<td>Beam-6dB</td>
</tr>
<tr>
<td>SpatialNR</td>
<td>Reference</td>
</tr>
<tr>
<td></td>
<td>Anchor</td>
</tr>
<tr>
<td></td>
<td>SpatialNR-0dB</td>
</tr>
<tr>
<td></td>
<td>SpatialNR-2dB</td>
</tr>
<tr>
<td></td>
<td>SpatialNR-4dB</td>
</tr>
<tr>
<td></td>
<td>SpatialNR-6dB</td>
</tr>
</tbody>
</table>

7.3 Results

The results are provided in separate sections for each outcome measure; SRT and MUSHRA sound quality.

7.3.1 SRT

The group mean SRTs and 95% confidence intervals for each of the four processing conditions at each of the three levels of microphone mismatch are shown in Figure 7-3. A two-way RM-ANOVA with processing condition and microphone mismatch as factors revealed a significant main effect of processing condition (F[3,39]=20.68, P<0.001), a significant main effect of microphone mismatch (F[2,26]=12.15, P<0.001) and a non-signification interaction term, processing x mismatch (F[6,77]=1.06, P=0.397).
7.3 Results

Intelligibility in 4-TB-full

![Intelligibility Graph]

Figure 7-3: Speech intelligibility in 4-TB-full as a function of microphone sensitivity mismatch. Group mean SRTs are shown and error bars show the 95% confidence interval.

Sound quality in 4-TB-full

![Sound Quality Graph]

Figure 7-4: MUSHRA sound quality ratings in 4-TB-full noise for various levels of microphone mismatch. Group mean data are shown with error bars indicating the 95% confidence interval.

The general trend that was observed with SRT results was that as microphone mismatch was increased, performance was reduced. Post-hoc comparisons of SRTs using the Student-Newman-Keuls method revealed many small but significant differences, indicating that
although the trends were small, they were highly consistent across the group. Given the non-significant interaction term (processing x mismatch) post-hoc comparisons were conducted by averaging the results for each main effect over the other factor. For the main effect of microphone mismatch, averaged across all processing conditions, the effect of changing the mismatch from 0 to 2 dB resulted in a 0.6 dB ($P=0.010$) reduction in SRTs, changing from 2 to 4 dB mismatch resulted in a further 0.4 dB ($P=0.042$) reduction in performance, and therefore an overall difference of 1 dB ($P<0.001$) SRT when the mismatch was changed from 0 to 4 dB. For the main effect of processing condition, averaged across all levels of microphone mismatch, there was a significant difference in SRTs between all paired conditions except the Standard/Beam pair. Worthy of note is that SpatialNR performed significantly better than all other processing conditions, 1.5 dB better than Standard ($P<0.001$), 1.4 dB better than Beam ($P<0.001$) and 0.5 dB better than Zoom ($P=0.039$).

Although the interaction term (processing x mismatch) was non-significant, a more detailed analysis of paired comparisons of processing conditions within each level of mismatch revealed interesting trends. For example, the effect of microphone mismatch in Standard and Beam conditions was not significant for any change in mismatch (all $P>0.18$), while for Zoom and SpatialNR there was a significant effect. For Zoom, there was a significant effect between 0 dB and 4 dB mismatch (1.4 dB SRT, $P=0.001$) and between 2 dB and 4 dB mismatch (0.8 dB SRT, $P=0.037$) but not 0 dB and 2 dB (0.6 dB SRT, $P=0.121$). For SpatialNR, there was a significant effect between 0 dB and 4 dB mismatch (1.5 dB SRT, $P<0.001$) and between 0 dB and 2 dB (0.8 dB SRT, $P=0.046$) but not between 2 dB and 4 dB mismatch (0.7 dB SRT, $P=0.068$).

With the exception of the Standard/Beam pair of processing conditions, the main effect of processing revealed all other pairs were significantly different. However, a post-hoc inspection within the microphone mismatch level of 4 dB revealed the effect of processing condition had disappeared and there was no significant difference between any pair of processing. This suggests that, at a mismatch of 4 dB, speech intelligibility was not significantly affected by the choice of noise reduction processing, contrary to the significant effect observed at both 0 dB and 2 dB mismatch.

### 7.3.2 MUSHRA

The initial analysis of MUSHRA results involved comparing the MUSHRA sound quality ratings for 0 dB mismatch that were reported in the baseline comparison, to those reported in the other comparisons. Paired t-tests (two-tailed) comparing the rating at 0 dB mismatch that was reported during the baseline comparison to the rating that was reported in each other comparison was made for each processing condition. The analysis showed that there was no significant difference for any of the processing conditions in ratings acquired in the two comparisons (Standard $P=0.299$, Zoom $P=0.432$, Beam $P=0.758$, SpatialNR $P=0.944$). This indicates that the absolute sound quality rating for each of the processing conditions at 0 dB mismatch was not likely influenced by the other conditions that were present during the comparison. That is, the ratings for each processing condition at 0 dB mismatch were consistent across the baseline comparison and the comparison made with other levels of mismatch.
Given the consistency in absolute ratings across different comparisons, the MUSHRA data was analysed with a two-way RM-ANOVA with processing condition and mismatch as the main factors. The data from the baseline comparison was not included in the two-way analysis. The data for the Reference and Anchor conditions was also excluded from the analysis. The two-way RM-ANOVA revealed a significant interaction between processing condition and mismatch \((F[9,117]=3.393, P<0.001)\), indicating the sound quality rating of different processing strategies was dependant on the level of microphone mismatch.

Student-Newman-Keuls post-hoc comparisons were conducted to analyse the interaction effect. For the Standard and Zoom conditions, the effect of microphone mismatch did not lead to a significant effect of microphone mismatch, even for the most extreme comparison between 0 dB and 6 dB mismatch. However, for Beam, the effect of mismatch was significant at 2 dB mismatch with a difference of 10.5 rating points compared to 0 dB mismatch \((P<0.001)\) and for SpatialNR the effect was significant at 4 dB mismatch with a difference of 7.5 rating points compared to 0 dB mismatch \((P=0.033)\). At 6 dB, the effect of mismatch continued to reduce performance such that the difference was 17.2 rating points for Beam \((P<0.001)\) and 11.2 rating points for SpatialNR \((P<0.001)\) compared to 0 dB mismatch.

A comparison between processing conditions at different levels of mismatch revealed that there was no significant difference in ratings between Standard and Zoom at any level of mismatch; they were relatively immune to sound quality changes over the range of mismatch tested \((P=0.899, 0.209, 0.131 \text{ and } 0.938 \text{ for mismatch of } 0, 2, 4 \text{ and } 6 \text{ dB respectively}). At 0 dB there was no significant difference between Standard, Zoom and Beam (all \(P>0.53)),\) and SpatialNR was rated significantly higher than all three. The difference in rating points was 10.1 compared to Standard \((P=0.006),\) 10.6 compared to Zoom \((P=0.012)\) and 12.9 compared to Beam \((P=0.004).\) As the mismatch was increased, the ratings for Beam and SpatialNR were reduced, while Standard and Zoom remained constant. Beam was rated with significantly poorer quality than all other conditions at mismatch levels of 2 dB and greater. In comparison to 0 dB mismatch, the effect of mismatch on Beam first became significant at 2 dB by a difference of 10.5 rating points \((P<0.001).\) While SpatialNR was rated significantly higher than all conditions at 0 dB mismatch, ratings of SpatialNR were significantly higher than only some conditions for greater mismatches; at 2 dB mismatch there was only a significant difference compared with Standard and Beam, by 11.8 \((P=0.005)\) and 22.1 \((P<0.001)\) rating points respectively, and at 4 dB and 6 dB mismatch, there was only a significant difference compared to Beam, by 17.4 \((P<0.001)\) and 18.9 \((P<0.001)\) rating points respectively.

### 7.4 Discussion and conclusion

The effect of microphone mismatch had an overall negative impact on speech intelligibility and sound quality for the group of CI users listening to speech in 4-TB-full noise, although the impact was different for the different algorithms that were evaluated. This trend was evident in the theoretical analysis, which showed increased error rates and decreased STI as the level of mismatch between the microphones was increased. At all levels of mismatch that were clinically evaluated, SpatialNR produced the best speech intelligibility and attracted the highest
sound quality ratings compared to Standard, Zoom and Beam. In terms of speech intelligibility, SpatialNR and Zoom were most affected by mismatch, and in terms of sound quality, SpatialNR and Beam were most affected. Once the mismatch was as extreme as 4 dB, differences in speech intelligibility and sound quality were not significant for all processing conditions (with the exception that the sound quality for Beam was significantly worse). That is, a mismatch of 4 dB obliterated any benefit obtained with any of the directional algorithms, such that no significant differences between any of the processing conditions were observed. Note the sound quality ratings for Beam were significantly worse than all other conditions at 2 dB mismatch and greater – this was probably due to target distortion that was introduced, which was not severe enough to impact speech understanding, but enough to impact speech quality. It appears that at a mismatch of 4 dB, it is not possible to achieve any directional benefit with the algorithms that were tested, at least in the 4-TB-full noise type used for evaluation, and possible to make things worse in the case of Beam.

The 4-TB-full noise type used is a difficult listening situation, and the maximum difference between any two processing conditions at 0 dB mismatch was less than 2 dB SRT and less than 13 sound quality rating points; nevertheless, there was a clear impact of microphone mismatch that was measured within the small range available. Evaluation in a different noise environment where the potential benefit of directional processing is greater (4-TB-rear for example) may expose a more (or less) dramatic roll-off in performance due to mismatch and/or differences between algorithms that were not highlighted in the present data.

The effect of microphone mismatch was detrimental and a 4 dB mismatch was enough to remove directional benefit altogether for all algorithms. Therefore, a recommendation on the absolute maximum tolerable mismatch between microphones is 4 dB or less. Speech intelligibility results suggest that by 2 dB there was a decrement in SpatialNR performance, but no decrement to other processing conditions until 4 dB. Speech quality data suggest that by 2 dB there was a decrement to Beam quality, but no decrement to SpatialNR until 4 dB. Therefore, the impact of microphone mismatch of less than 2 dB may not significantly decrease performance for CI users in their everyday lives. Given the discrete levels of mismatch that were evaluated, a recommendation for maximum tolerable mismatch is 2 dB.

Real microphones may degrade differently to the broad-band sensitivity changes that were modelled in this study. For example, high frequency sensitivity of one microphone may be affected by a dirty microphone cover. In this situation, the translation of mismatch to performance is more complex, involving the spectral content of the target and noise signal. Additionally, the phase mismatch may also be affected, resulting in degraded directional performance, although phase mismatch was not investigated in the present study. Nevertheless, the model of broad-band mismatch provides an indication of how mismatch relates to performance under various directional processing conditions, and confirms that the degree of mismatch should be less than 4 dB in order to obtain benefit from these algorithms. Some caution needs to be exercised in generalising these results to different acoustic situations since the conclusions drawn pertain to 4-TB-full noise in low reverberant conditions.
Chapter 8

Experiment V

BTE device take-home evaluation

This final experiment utilised a BTE sound processor implementation in order to facilitate evaluation outside the laboratory. The objective was to demonstrate user preference of different sound processing conditions and parameter settings in the real-world. The maximum attenuation parameter of the SpatialNR algorithm was varied and tested. The experiment contained a broad laboratory-based evaluation involving speech intelligibility, sound quality and acceptable noise level tests across four different noise types (2-SWN-rear, 4-TB-rear, 4-TB-full, 20-TB-full).

8.1 Introduction

Experiment V was designed to evaluate the SpatialNR algorithm outside the laboratory so that CI subjects could listen to the SpatialNR algorithm during normal everyday use. To achieve this, the SpatialNR algorithm was implemented on a research version of the CP900 BTE processor. Subjects were asked to compare different listening programs and vote for their preferred program using the CR230 remote control device used to control the sound processor. In this way, direct feedback regarding the subjects’ preferences were obtained in the field at the time subjects made their assessment. The device automatically analysed the characteristics of the acoustic environment and performed scene classification, which was also stored on the CR230 with the accompanying vote data.

Laboratory-based assessment during this experiment concentrated on evaluation in a variety of different noise types, whilst the maximum attenuation of the SpatialNR algorithm was systematically varied to understand the impact on performance in these conditions. CI subjects undertook several different listening tasks in addition to speech intelligibility testing using SRTs; sound quality was assessed using the MUSRHA task and acceptable noise level (ANL) was
also measured as it was thought to contribute to subjective benefit that CI users report in the field (Donaldson et al., 2009). During the take-home evaluation, subjects gained experience listening to the SpatialNR processing, and acclimatisation was evaluated through SRT testing in 4-TB-rear noise at each visit. Although it was not expected that long term exposure to SpatialNR would lead to intelligibility changes over time, the current study provided a platform to investigate the possibility.

In order to compare the results of the voting task with more a conventional subjective feedback metric, a modified SSQ questionnaire was administered as part of the take-home evaluation. It was intended that the voting task would provide a more accurate representation of listening preference since it allowed for direct evaluation in the field, whereas the questionnaire required subjects to recall different listening situations and their program preference in those situations.

Additionally, as a benchmarking exercise, a group of normal hearing subjects were evaluated unaided in three of the four noise conditions used for laboratory-based testing. Whilst it was expected that normal hearing SRT performance would generally be better than CI across all of these noise types, it was expected that the introduction of noise reduction processing would improve CI performance, closing the gap in performance between the two groups.

### 8.2 Methods

In this experiment, CI subjects undertook both laboratory-based and field-based tasks. The primary goal was to understand how the maximum attenuation parameter of the SpatialNR algorithm affected the laboratory-based metrics (speech intelligibility, sound quality and acceptable noise level) and how the algorithm was accepted in the field.

#### 8.2.1 Processing conditions

All testing was performed using the CP900 SpatialNR implementation (BTE), whilst selected laboratory tests retained the xPC implementation used in all previous experiments. This was done for the purpose of comparing the two platforms. Processing conditions included Standard, Zoom, Beam(Z) and SpatialNR. Note that Beam(Z) used Zoom directionality as the speech reference signal for the adaptive noise cancellation stage. This distinguishes it from Beam used in Experiments I-IV which used Standard directionality as the speech reference for the adaptive noise cancellation stage. The reason for using Zoom as the speech reference was to enable subjects to make comparisons between Beam and SpatialNR such that both algorithms used to the same speech reference signal, avoiding switching artefacts.
Maximum attenuation parameter

The SpatialNR parameter under investigation in Experiment V was the maximum attenuation parameter, $\gamma$. Its function was to switch the SpatialNR algorithm on and off and, more importantly, to provide a smooth mechanism for doing so. To do this, the estimated clean signal output from SpatialNR, $\hat{X}_k[n]$, was mixed with the noisy input signal, $S_k[n]$, under the control of the maximum attenuation parameter, $\gamma$, to produce the output signal $Y_k[n]$,

$$Y_k[n] = \gamma \hat{X}_k[n] + (1 - \gamma)S_k[n] \quad (8-1)$$

The parameter value range was from 0 (or -inf dB), with no limit on the maximum attenuation and therefore full noise reduction, to 1 (or 0 dB) with full limit on the maximum attenuation and therefore no noise reduction. In this evaluation, maximum attenuation values of -96, -14, -7 and 0 dB were evaluated (Table 8-1). When the maximum attenuation is 0 dB, there is no noise reduction gain applied; therefore, the algorithm output reverts to the speech reference signal, Zoom.

**SNR-NR**

The Nucleus 6 CP900 sound processor had a single channel noise reduction algorithm called SNR-NR (Mauger et al., 2014), which was evaluated in combination with SpatialNR in this study. SNR-NR was previously evaluated in combination with various directionality settings and found to provide speech intelligibility benefit in stationary noise, but not in modulated competing babble (Dawson et al., 2011; Hersbach et al., 2012; Mauger et al., 2014). Given the established benefit of SNR-NR, SNR-NR and SpatialNR were not directly compared against each in the present study, however they were tested in combination. The SNR-NR algorithm was enabled to create the condition SpatialNR(MA14)NR with the purpose of determining any additional benefit that might be provided by the SNR-NR algorithm in addition to SpatialNR.

**Implementation platform**

In order to verify the full implementation on CP900, both xPC and BTE implementations of Zoom, SpatialNR(MA14) and SpatialNR(MA96) were compared in 4-TB-rear noise. This noise
type has previously demonstrated the largest improvements with SpatialNR, and was therefore expected to be the noise type most sensitive to any differences in implementation. Furthermore, SpatialNR(MA14)NR was compared across the two implementation platforms in 2-SWN-rear, since it was the most sensitive noise type for the NR algorithm. The performance with the two implementation platforms was expected to be equal in all cases, and the comparisons were planned such that underlying differences in individual algorithms could be determined.

8.2.2 Laboratory-based evaluation

In the laboratory, evaluation took place in the Whitford sound treated booth. Four different noise conditions were used for testing: 2-SWN-rear, 4-TB-rear, 4-TB-full and 20-TB-full.

Laboratory-based outcome metrics involving speech intelligibility (SRT), sound quality (MUSHRA) and acceptable noise level (ANL) were all evaluated in the four different noise types.

The entire list of processing conditions evaluated in the laboratory is shown in Table 8-2. For the purpose of analysis, the conditions were grouped in order to target specific research questions:

- **Overall Performance:** Standard, Zoom, Beam(Z), SpatialNR(MA14)
- **Effect of MA:** Zoom, SpatialNR(MA7), SpatialNR(MA14), SpatialNR(MA96)
- **Effect of SNR-NR:** SpatialNR(MA14), SpatialNR(MA14)NR
- **Effect of acclimatisation:** Zoom, SpatialNR(MA14)
- **Effect of platform:** Zoom, SpatialNR(MA14), SpatialNR(MA96), SpatialNR(MA14)NR

8.2.3 Field-based evaluation

Subjective, field-based outcome metrics were Voting and SSQ questionnaire. There were five different configurations used for the field-based assessment, each involving the comparison of two or, in the case of the final comparison, three listening programs. The voting procedure involved subjects changing listening programs using the remote control device (CR230) and pressing a ‘vote’ button to indicate the program that was considered most helpful in each listening situation. Full details of the procedure can be found in Methods Chapter 3.

The modified SSQ questionnaire with two additional questions on overall preference was also administered, which provided subjective feedback on the two listening programs. Overall program preference and degree was collected for two situations: quiet and noise. Degree of preference was converted to a preference score on a scale from -3 to +3. Negative values indicated a preference for the baseline condition in the comparison, whilst positive values indicated a preference for the experimental condition (SpatialNR or SNR-NR). The questionnaire was not administered for the final comparison, which involved three programs as the modified questionnaire only supported the rating of two programs.
Table 8-2: Processing conditions used for evaluation, showing details of sound processing, implementation platform and laboratory-based tests that were involved.

<table>
<thead>
<tr>
<th>Processing condition</th>
<th>CP900 directionality</th>
<th>Experimental algorithm</th>
<th>Platform</th>
<th>Noise types used for laboratory-based tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard</td>
<td>Standard</td>
<td>-</td>
<td>BTE</td>
<td>SRT MUSHRA ANL All All All</td>
</tr>
<tr>
<td>Zoom</td>
<td>Zoom</td>
<td>-</td>
<td>BTE</td>
<td>All All All 4-TB-rear</td>
</tr>
<tr>
<td>Beam(Z)</td>
<td>Beam with Zoom speech reference</td>
<td>-</td>
<td>BTE</td>
<td>All All All</td>
</tr>
<tr>
<td>SpatialNR(MA96)</td>
<td>Zoom</td>
<td>SpatialNR MA = -96 dB</td>
<td>BTE</td>
<td>All All All 4-TB-rear</td>
</tr>
<tr>
<td></td>
<td></td>
<td>xPC</td>
<td>xPC</td>
<td>- -</td>
</tr>
<tr>
<td>SpatialNR(MA14)</td>
<td>Zoom</td>
<td>SpatialNR MA = -14 dB</td>
<td>BTE</td>
<td>All All All 4-TB-rear</td>
</tr>
<tr>
<td></td>
<td></td>
<td>xPC</td>
<td>xPC</td>
<td>- -</td>
</tr>
<tr>
<td>SpatialNR(MA7)</td>
<td>Zoom</td>
<td>SpatialNR MA = -7 dB</td>
<td>BTE</td>
<td>All All All</td>
</tr>
<tr>
<td>SpatialNR(MA14)NR</td>
<td>Zoom</td>
<td>SpatialNR MA = -14 dB</td>
<td>BTE</td>
<td>All All All</td>
</tr>
<tr>
<td></td>
<td></td>
<td>xPC</td>
<td>xPC</td>
<td>2-SWN-rear - -</td>
</tr>
</tbody>
</table>

The take home voting task involved the comparison of two, or in the case of the final comparison, three test programs. In the first three comparisons, the SpatialNR algorithm was evaluated, in the fourth comparison SNR-NR was evaluated and in the last comparison directionality alone was evaluated. The CP900 SCAN feature was utilised to record the sound class in order to categorise each vote event. The SCAN feature also provided data relating to the amount time spent listening to each program and in each sound class. In addition to providing a sound classification, the SCAN feature also allowed automatic switching of the underlying directionality of the system depending on the class, and modified SCAN programs were generated in order to make the appropriate comparisons (Table 8-3).
Table 8-3: Configuration of microphone directionality for modified SCAN programs used for field-based assessment. SCAN was the CP900 default setting not used in this experiment, SCAN1 was used as the base program (SpatialNR disabled), and SCAN2 was used as the test program with SpatialNR enabled.

<table>
<thead>
<tr>
<th>Class</th>
<th>Quiet</th>
<th>Speech</th>
<th>Speech in noise</th>
<th>Noise</th>
<th>Music</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCAN</td>
<td>Standard</td>
<td>Standard</td>
<td>Standard + Beam</td>
<td>Zoom</td>
<td>Standard</td>
</tr>
<tr>
<td>SCAN1</td>
<td>Standard</td>
<td>Standard</td>
<td>Zoom + Beam</td>
<td>Zoom</td>
<td>Standard</td>
</tr>
<tr>
<td>SCAN2</td>
<td>Standard</td>
<td>Standard</td>
<td>Zoom</td>
<td>Zoom</td>
<td>Standard</td>
</tr>
</tbody>
</table>

Zoom-SpatialNR(MA14)

In the first comparison called Zoom-SpatialNR(MA14), the two programs were Zoom and Zoom + SpatialNR(MA14). The SCAN feature was configured so that the underlying directionality was always Zoom, so did not change automatically as the class changed, providing a stable comparison across all sound classes. When the user changed programs, the only effect was to introduce SpatialNR.

SCAN-SpatialNR(MA14)

In the second comparison called SCAN-SpatialNR(MA14), the two programs were SCAN1 and SCAN2 + SpatialNR(MA14). SCAN1 and SCAN2 were both modified versions of SCAN (Table 8-3), where the underlying control of directionality was slightly modified from the CP900 default operation. The Speech in Noise class was the only modification made, so that Zoom + Beam was compared with Zoom + SpatialNR in that class. In each of the remaining sound classes, the only difference between programs was the addition of SpatialNR. During the field-based assessment, the SCAN feature changed the underlying directionality according to modified SCAN configuration, and the user effectively compared the two programs as shown in Table 8-4 in each sound class.

SCAN-SpatialNR(MA7)

In the third comparison called SCAN-SpatialNR(MA7), the two programs were SCAN1 and SCAN2 + SpatialNR(MA7). The comparison was similar to the previous SCAN-SpatialNR(MA14) comparison, instead using a less aggressive setting for the SpatialNR maximum attenuation (MA) parameter. This comparison was reserved only for subjects who exhibited an overall dislike for the more aggressive setting. The purpose of the less aggressive comparison was to search for an acceptable MA setting for those subjects who did not accept the stronger setting.
Table 8-4: Program configurations used for the field-based assessment. With the exception of the ‘Directionality’ comparison, the underlying directionality in each sound class was consistent across the two programs being evaluated, such that only the experimental algorithm varied.

<table>
<thead>
<tr>
<th>Comparison Name</th>
<th>Program Name</th>
<th>Quiet</th>
<th>Speech</th>
<th>Speech in noise</th>
<th>Noise</th>
<th>Music</th>
</tr>
</thead>
<tbody>
<tr>
<td>SpatialNR(MA14)</td>
<td>Zoom</td>
<td>Zoom</td>
<td>Zoom</td>
<td>Zoom</td>
<td>Zoom</td>
<td>Zoom</td>
</tr>
<tr>
<td>SpatialNR(MA14)</td>
<td>Zoom + SpatialNR (MA14)</td>
<td>Zoom + SpatialNR (MA14)</td>
<td>Zoom + SpatialNR (MA14)</td>
<td>Zoom + SpatialNR (MA14)</td>
<td>Zoom + SpatialNR (MA14)</td>
<td></td>
</tr>
<tr>
<td>SCAN-SpatialNR(MA14)</td>
<td>SCAN1</td>
<td>Standard</td>
<td>Standard</td>
<td>Zoom + Beam</td>
<td>Zoom</td>
<td>Standard</td>
</tr>
<tr>
<td>SCAN2 + SpatialNR(MA14)</td>
<td>Standard + SpatialNR (MA14)</td>
<td>Standard + SpatialNR (MA14)</td>
<td>Zoom + SpatialNR (MA14)</td>
<td>Zoom + SpatialNR (MA14)</td>
<td>Standard + SpatialNR (MA14)</td>
<td></td>
</tr>
<tr>
<td>SCAN-SpatialNR(MA7)</td>
<td>SCAN1</td>
<td>Standard</td>
<td>Standard</td>
<td>Zoom + Beam</td>
<td>Zoom</td>
<td>Standard</td>
</tr>
<tr>
<td>SCAN2 + SpatialNR(MA7)</td>
<td>Standard + SpatialNR (MA7)</td>
<td>Standard + SpatialNR (MA7)</td>
<td>Zoom + SpatialNR (MA7)</td>
<td>Zoom + SpatialNR (MA7)</td>
<td>Standard + SpatialNR (MA7)</td>
<td></td>
</tr>
<tr>
<td>SCAN-NR</td>
<td>SCAN1</td>
<td>Standard</td>
<td>Standard</td>
<td>Zoom + Beam</td>
<td>Zoom</td>
<td>Standard</td>
</tr>
<tr>
<td>Zoom</td>
<td>Zoom</td>
<td>Zoom</td>
<td>Zoom</td>
<td>Zoom</td>
<td>Zoom</td>
<td></td>
</tr>
<tr>
<td>Beam(Z)</td>
<td>Zoom + Beam</td>
<td>Zoom + Beam</td>
<td>Zoom + Beam</td>
<td>Zoom + Beam</td>
<td>Zoom + Beam</td>
<td></td>
</tr>
</tbody>
</table>

**SCAN-NR**

In the fourth comparison called **SCAN-NR**, the two programs were SCAN1 and SCAN1+NR. The assessment was designed to evaluate SNR-NR using the same protocol as was used to evaluate SpatialNR. The comparison did not involve SpatialNR at all. The underlying directionality was automatically adjusted according to the SCAN1 configuration, therefore only the NR algorithm changed with program. The purpose of the comparison was to provide an anchor for the voting results since the voting paradigm was new and had not been previously validated. Field-
and laboratory-based assessments of SNR-NR in past studies had been positive (Hersbach et al., 2012), and as such, voting data in favour of SNR-NR was expected.

**Directionality**

In the final comparison called **Directionality**, three programs were used which had fixed directionality: Standard, Zoom, and Beam(Z). Although the SCAN feature was used to record the sound class, this comparison did not use SCAN to change directionality. Therefore, the directionality was fixed in each class and changed only when the user changed programs. The purpose of this comparison was to find the preferred directionality setting in each sound class, as determined by the voting data.

**8.2.4 Test protocol**

Subjects visited the laboratory on a total of five occasions, separated by two week periods between each visit. The testing during different visits was organised around the noise type used for evaluation, such that each of the four noise types was evaluated in a different visit, and the order of testing of the different noise types was randomised. This formed the core component of the testing, such that at each visit, SRT, MUSHRA and ANL tests were conducted in the same noise type, and the noise type changed as the visits progressed. The processing conditions that were evaluated were consistent across the different noise types except for the 4-TB-rear noise where xPC conditions were added to compare implementation platforms. A single SRT run was collected for each processing condition throughout the experiment. The order of testing of processing conditions was randomised within session.

The core testing occurred at visits 2-5. In addition, on the first and all subsequent visits at the beginning of the test session, Zoom and SpatialNR(MA14) were evaluated in 4-TB-rear noise, which provided data to test for any acclimatisation over the experimental period. The order of testing of these two conditions was randomised. In the visit that was dedicated to the 4-TB-rear, these two conditions were already tested as part of the core testing, and therefore not additionally tested at the beginning of the session.

Interleaved with the laboratory testing was the take-home component of the evaluation, which occurred between visits. During each two week period, subjects were provided with two listening programs and asked to perform the voting task. Subjects made four compulsory comparisons and a possible fifth comparison that depended on the subject’s overall preference in the SCAN-SpatialNR(MA14) comparison, determined by the questionnaire-based responses to overall preference in quiet and noisy conditions. If the preference for the baseline (SCAN) condition was “Much prefer” or “Moderately prefer” in quiet or “Much prefer”, “Moderately prefer”, “Slightly prefer” or “No difference” in noise then the SCAN-SpatialNR(MA7) comparison was administered, otherwise it was skipped for that subject. All subjects performed the comparisons in the same prescribed order, resulting in a common exposure to the SpatialNR algorithm across the group.
8.3 Laboratory-based results

A group of 20 normal hearing subjects took part in the benchmarking evaluation, in which unaided SRTs were collected in three of the four noise types. Their mean age was 23 years 3 months and all had pure tone thresholds of 20 dB HL or better at octave frequencies over the range 250 - 8000 Hz. No attempt was made to age-match the NH group to the CI group. The NH group were evaluated unaided using their natural normal hearing. A single test session was conducted, in which two SRTs were collected in each noise type, 4-TB-rear, 4-TB-full and 20-TB-full. For each subject, the noise types were tested in random order, then tested in reverse order to provide counter-balancing. SRTs were averaged across the two SRT runs for each noise type. Evaluation took place in the Whitford sound treated booth.

8.3.1 Overall performance

Overall performance was assessed by comparing SpatialNR to the baseline processing conditions, Standard, Zoom and Beam(Z).

Figure 8-1 shows the group mean SRT, MUSHRA, and ANL data for the BTE implementations of Standard, Zoom, Beam(Z) and SpatialNR(MA14), which summarises the degree of noise reduction obtained in the four different noise types. Across the three objective measures, there was an overall trend which suggests superior performance of the SpatialNR algorithm. For statistical analyses, separate RM-ANOVAs were performed on each of the outcome measures using processing condition (Standard, Zoom, Beam(Z) and SpatialNR(MA14)) and noise type (2-SWN-rear, 4-TB-rear, 4-TB-full, and 20-TB-full) as main factors.

SRT

A two-way RM-ANOVA on SRTs revealed a significant interaction between the main factors (F[9,126]=10.70, P<0.001) indicating that the benefit of processing was dependent on the noise type used for evaluation. Post-hoc analysis using the Student-Newman-Keuls method follows, with asterisks used in to denote significant comparisons.

In 2-SWN-rear, there was no significant difference between Beam(Z) and SpatialNR(MA14) (P=0.198), but both were significantly better than Zoom by 2.5 dB (P<0.001) and 1.7 dB (P=0.002) respectively, and Zoom was significantly better than Standard by 3.0 dB (P<0.001).
Figure 8-1: Group SRT, MUSHRA and ANL results for BTE implementation of the four main directional algorithms (Standard, Zoom, Beam(Z) and SpatialNR(MA14)) that were evaluated, separated by noise type. Error bars show 95% confidence interval. Post-hoc comparisons between processing conditions that were statistically significant are marked with asterisks (*P<0.05, **P<0.01 and ***P<0.001).

In 4-TB-rear, there was a significant difference between all conditions, with SpatialNR(MA14) providing the best performance in this noise type. SpatialNR(MA14) was 2.2 dB better than Beam(Z) (P< 0.001), which was better than Zoom by 1.5 dB (P=0.002), which was better than
Standard by 3.4 dB (P<0.001). The benefit of SpatialNR(MA14) over Standard was altogether 7.2 dB (P<0.001).

In 4-TB-full, there was no significant difference between SpatialNR(MA14), Beam(Z), and Zoom, but all were significantly better than Standard by 2.7 dB (P<0.001), 2.8 dB (P<0.001) and 2.6 dB (P<0.001), respectively.

In 20-TB-full, the pattern of results was similar to 4-TB-full, such that there was no significant difference between SpatialNR(MA14), Beam(Z), and Zoom, but all were significantly better than Standard by 1.7 dB (P<0.002), 2.1 dB (P<0.001) and 1.5 dB (P<0.002) respectively.

**MUSHRA**

A two-way RM-ANOVA of MUSHRA scores was performed with the responses to the hidden reference condition removed from the analysis. A significant main effect of noise type was found (F[3,42]=5.109, P=0.004), a significant main effect of processing condition was found (F[3,42]=23.663, P<0.001), and the interaction term was not significant (F[9,126]=0.766, P=0.648). When averaged across all noise types, post-hoc comparisons revealed that SpatialNR(MA14) was rated as significantly higher quality than all other processing conditions: 10.2 points higher than Standard (P<0.001), 3.2 points higher than Zoom (P=0.013), and 4.0 points higher than Beam(Z) (P=0.007). Both Beam(Z) and Zoom were rated significantly higher quality than Standard (6.18 points P<0.001 and 7.0 points P<0.001 respectively), but not significantly different from each other (P=0.535).

**ANL**

A two-way RM-ANOVA on ANLs revealed the main effect of noise type was not significant (F[3,42]=2.507, P=0.072), the main effect of processing condition was significant (F[3,42]=2.925, P=0.045), and the interaction term was not significant (F[9,126]=0.669, P=0.735). When averaged across all noise types, post-hoc comparisons revealed the only significant comparison was between Standard and SpatialNR(MA14) where the difference was 2.5 dB (P=0.031).

**8.3.2 Effect of MA**

The effect of MA was assessed by evaluating at parameter values of 0, -7, -14 and -96 dB.

SRTs, MUSHRA sound quality ratings and ANLs were collected for maximum attenuation (MA) values of 0 (Zoom), -7 (MA7), -14 (MA14) and -96 dB (MA96) in all four noise types shown in Figure 8-2. Note that the zoom condition is equivalently SpatialNR(MA0). For all outcome measures, scores have been plotted as the difference relative to Zoom, and therefore represent the benefit compared to Zoom for each MA setting. Separate two-way RM-ANOVA were performed on each outcome measure using MA (0, -7, -14, and -96 dB) and noise type (2-SWN-rear, 4-TB-rear, 4-TB-full, and 20-TB-full) as main factors.
Figure 8-2: The benefit of maximum attenuation on speech intelligibility (SRT), sound quality (MUSHRA) and acceptable noise level (ANL). Mean benefit compared to Zoom condition in each of the four noise types is shown. Error bars show 95% confidence intervals. Post-hoc comparisons that were statistically significant are marked with asterisks (*P<0.05, **P<0.01 and ***P<0.001).

SRT

The two-way RM-ANOVA on SRTs revealed a significant interaction between the main factors (F[9,126]=6.671, P<0.001) suggesting that the effect of MA varied with noise type. Post-hoc comparisons within each noise type revealed significant differences due to MA in the two rear
8.3 Laboratory-based results

half noises (2-SWN-rear, 4-TB-rear), and no significant differences due to MA in the full circle noises (4-TB-full, 20-TB-full).

In 2-SWN-rear, performance significantly improved from MA0 to MA7 by 1.5 dB (P=0.002), but did not change significantly as MA was further changed to MA14 or MA96. However, the best performance was obtained with MA96 which was altogether 2.0 dB better than MA0 (P<0.001).

In 4-TB-rear, performance was significantly improved from MA0 to MA7 by 2.5 dB (P<0.001) and further improved at MA14 by another 1.3 dB (P=0.007). Performance did not significantly improve further at MA96, although this is where the best performance was obtained, altogether 4.2 dB better than MA0 (P<0.001).

In 4-TB-full and 20-TB-full, there were no significant differences between any conditions. That is, there was no significant effect of changing the MA value in these two full circle noise types.

MUSHRA

The two-way RM-ANOVA on MUSHRA scores with the Reference condition removed from the analysis revealed a significant effect of noise type (F[3,42]=6.104, P=0.002), a significant effect of MA (F[3,42]=4.391, P=0.009) and an interaction term that was not significant (F[9,126]=0.948, P=0.486). Averaged across all noise types, all MA settings received significantly higher sound quality ratings than Zoom. MA96 received the highest rating, 4.2 points better than Zoom (P=0.007), MA14 was 3.2 points better than Zoom (P=0.028) and MA7 was 2.5 points better than Zoom (P=0.042). There were no significant differences between MA7, MA14 and MA96.

ANL

The two-way RM-ANOVA on ANLs revealed a significant effect of noise type (F[3,42]=3.016, P=0.040), a significant main effect of MA (F[3,42]=3.740, P=0.018) and an interaction term that was not significant (F[9,126]=0.416, P=0.924). Averaged across all noise types, only the comparison between SpatialNR(MA96) and Zoom revealed a significant difference (2.0 dB, P=0.013).

8.3.3 Effect of SNR-NR

The single-channel SNR-NR algorithm (NR) was evaluated in combination with SpatialNR(MA14). The benefit of NR was assessed through the comparison of two conditions, which were configured identically while NR was enabled or disabled.

The SRT, MUSHRA and ANL results under the two conditions SpatialNR(MA14) and SpatialNR(MA14)NR are shown in . For each outcome measure, a separate two-way RM-ANOVA was performed with NR (off, on) and noise type (2-SWN-rear, 4-TB-rear, 4-TB-full, 20-TB-full) as main factors.
Figure 8-3: Benefit in speech intelligibility (SRT), sound quality (MUSHRA) and acceptable noise level (ANL) of single channel noise reduction (NR) evaluated in combination with SpatialNR(MA14) in each of the four noise types. The benefit of NR on relative to NR off is shown, error bars show 95% confidence intervals. Post-hoc comparisons that were statistically significant are marked with asterisks (*P<0.05, **P<0.01 and ***P<0.001).

The two-way RM-ANOVA analyses revealed there was no significant effect of the NR on SRTs (F[1,14]=0.672, P=0.426), on MUSHRA sound quality ratings (F[1,14]=0.002, P=0.963) or ANLs (F[1,14]=3.926, P=0.068). Post-hoc analyses (Student-Newman-Keuls) of results in 2-SWN-rear were performed due to the expectation that NR was expected to provide some benefit in that noise type. The analysis revealed that the 0.9 dB SRT effect of NR in 2-SWN-rear
was not quite significant (P=0.057), there was no significant effect on MUSHRA sound quality ratings (P=0.679), and there was a significant benefit in ANL of 3.6 dB (P=0.004).

### 8.3.4 Effect of acclimatisation

Over the course of the experiment, subjects were exposed to the SpatialNR algorithm during their take home use of the research version of the CP900 sound processor. In this section, the effect of acclimatisation to the algorithm is assessed.

The performance of Zoom and SpatialNR(MA14) was measured in 4-TB-rear at each of five sequential visits to the laboratory and the SRT results are shown in Figure 8-4. The benefit of SpatialNR(MA14) over Zoom was calculated at each visit and is presented in Figure 8-5. A two-way RM-ANOVA on the SRTs with processing condition (Zoom, SpatialNR(MA14)) and session (1-5) as main factors revealed a significant effect of processing condition (F[1,14]=162.6, P<0.001). This was expected due to the difference between Zoom and SpatialNR. There was no significant main effect of session (F[4,56]=0.560, P=0.692) and no significant interaction between the main factors (F[4,56]=0.483, P=0.748). Further to the analysis in order to assess any effect of acclimatisation, a one-way RM-ANOVA was performed on the benefit of SpatialNR over Zoom (i.e. on the difference scores), which revealed no significant effect of session (F[4,56]=0.483, P=0.748). That is, the benefit of SpatialNR did not change with test session.

### 8.3.5 Effect of implementation platform

Four different sound processing configurations, Zoom, SpatialNR(MA14), SpatialNR(MA96), and SpatialNR(MA14)NR were evaluated on both implementation platforms, BTE and xPC, via speech intelligibility (SRT), with the results shown in Figure 8-6. A paired t-test was performed on each pair of processing conditions that were identical apart from the implementation platform. There was no significant difference found between any pair of conditions. Zoom in 4-TB-rear (P=0.382), SpatialNR(MA14) in 4-TB-rear (P=0.160), SpatialNR(MA96) in 4-TB-rear (P=0.794), and SpatialNR(MA14)NR in 2-SWN-rear (P=0.290).

### 8.3.6 Normal hearing benchmark

The mean SRT and 95% confidence interval for the normal hearing (NH) group are shown in Figure 8-7 for three of the four noise types. SRTs were not collected for NH in the 2-SWN-rear noise type. The mean SRT’s were -11.0, -10.2, and -7.1 dB for 4-TB-rear, 4-TB-full and 20-TB-full noise types respectively. In order to compare CI and NH performance, also shown in Figure 8-7 are the group average SRTs for CI recipients listening to Standard and SpatialNR(MA14) processing. This was done in order to represent baseline processing (Standard) and the most effective noise reduction (SpatialNR(MA14)).
Performance over time in 4-TB-rear

![Graph](image1)

**Figure 8-4:** The effect of exposure to Zoom and SpatialNR(MA14) on speech intelligibility outcomes. SRTs in 4-TB-rear noise are presented with error bars showing 95% confidence intervals.

Performance benefit of SpatialNR over time in 4-TB-rear

![Graph](image2)

**Figure 8-5:** The effect of exposure to SpatialNR on speech intelligibility outcomes presented as the difference between Zoom and SpatialNR(MA14). Error bars show 95% confidence intervals.
Figure 8-6: Effect of implementation platform on speech intelligibility. SRT’s are shown for four pairs of conditions (connected with a line) that were identical apart from the implementation platform. Error bars show 95% confidence intervals.

Figure 8-7: Unaided normal hearing (NH) SRTs in comparison to CI performance using both Standard directionality and with SpatialNR(MA14). NH were evaluated in three of the four different noise types. Data points show the mean and 95% confidence intervals.
A 2 (hearing modality) by 3 (noise type) mixed ANOVA with hearing modality (NH, CI) as the between-subjects factor and noise type (4-TB-rear, 4-TB-full, 20-TB-full) as the within-subjects factor revealed a significant interaction between hearing modality and noise type (F[2,66]=92.073, P<0.001). This indicates that while the performance of the two groups were markedly different, where the NH group clearly outperformed the CI group by 14.9 dB in 4-TB-rear (P<0.001), 15.0 dB in 4-TB-full (P<0.001) and 9.3 dB in 20-TB-full (P<0.001), the two groups were differently affected by the noise types. CI performance was 2.7 dB worse in 4-TB-full compared to CI performance in 20-TB-full (P<0.001), while NH performance went in the opposite direction, improving by 3.1 dB in 4-TB-full compared to 20-TB-full (P<0.001). The difference between 4-TB-rear and 4-TB-full was not significant for NH group (P=0.125) or CI group (P=0.068).

The mixed ANOVA revealed the NH group outperformed the CI group listening to SpatialNR(MA14) processing by 8.7 dB in 4-TB-rear (P<0.001), 12.4 dB in 4-TB-full (P<0.001) and 7.6 dB in 20-TB-full (P<0.001).

### 8.4 Field-based results

The voting data and questionnaire data for the field-based assessment are presented in this section. A discussion of the field-based results follows, which concentrates on an analysis of the voting data. The analysis includes specific sections for each of the voting comparisons and a discussion on general voting patterns and observations. Relationships between voting preference and laboratory-based outcomes are analysed, so too are the relationships between voting and questionnaire preferences.

#### 8.4.1 Individual voting preference

Normalised vote scores for individual subjects for the first four comparisons are shown in Figure 8-8. The location of the bubble on the x-axis indicates the normalised vote score, the size of the bubble indicates the number of votes that contributed to that data point and colour indicates the sound class to which the votes were allocated. Data points to the left of 0 on the x-axis represent a majority of votes in favour of the baseline condition (either Zoom or SCAN, depending on the comparison), while data points to the right are majority in favour of the experimental program (either SpatialNR or SNR-NR). The Directionality comparison involved three programs, so the proportion of votes for each program were calculated and plotted directly on a three-way axis in Figure 8-9. Each bubble represents votes from an individual subject, where the location of the bubble indicates the proportion of votes for each of the three programs, the size of the bubble indicates the number of votes and the colour of the bubble indicated the sound class.
Figure 8-8: Individual raw voting data for comparisons (A) Zoom-SpatialNR(MA14), (B) SCAN-SpatialNR(MA14), (C) SCAN-SpatialNR(MA7) and (D) SCAN-NR. Size of bubble indicates the number of votes while colour indicates the sound class in which votes were made. Data are plotted on the x-axis as the normalised vote score such that the zero line indicates equal votes in each program.
Figure 8-9: Distribution of votes for the directionality comparison. The location of the data points indicate the proportion of votes in each of the three programs, where the tips of the triangle plot indicate all votes for that program. The size of the bubble indicates the number of votes and the colour indicates the sound class.

For the purpose of comparing voting patterns across different sound classes, the association of individual subject voting scores across sound class are shown in Figure 8-10 for the Zoom-SpatialNR(MA14), SCAN-SpatialNR(MA14) and SCAN-NR comparisons. The SCAN-Spatial(MA7) and Directionality comparisons are not shown due to the low number of votes outside the Quiet class for those comparisons. A linear model with one parameter was fit to the data of the form $y = ax$. The parameter, $a$, was estimated using least squares regression. The solid lines in Figure 8-10 show the model estimate while the dotted lines indicate the 95% C.I. for the estimate of the model parameter, $a$. A negative gradient of the 95% C.I. indicates the model parameter was not significantly different from zero. Positive gradients on both the lower and upper 95% C.I. indicate the parameter estimate was significantly different from zero, and hence a significant association between the voting scores in the two different sound classes.
Figure 8-10: Individual vote scores for the Zoom-SpatialNR(MA14), SCAN-SpatialNR(MA14) and SCAN-NR comparisons. Plots show association of vote scores across different sound classes. Solid lines represent the model $y=ax$. Dotted lines show the 95% C.I. for the estimate of the model parameter, $a$. A positive gradient on both the upper and lower 95% C.I. indicates the model fit was significant.
Figure 8-11: Group summary of voting data for the first four voting comparisons (A) Zoom-SpatialNR(MA14), (B) SCAN-SpatialNR(MA14), (C) SCAN-SpatialNR(MA7) and (D) SCAN-NR. Median vote scores for the three dominant sound classes are shown with p-values for one-sample signed rank tests against a hypothesised median score of 0 (*p<0.05, **p<0.01 and ***p<0.001). The table indicates the number of subjects that contribute to the median scores, divided by overall preference into categories against/equal/for SpatialNR/SNR-NR. The mean number of vote events per subject and scene are also shown.
8.4 Field-based results

![Distribution of sound class independent vote score for the first four voting comparisons. Box and whisker plots show the 10, 25, 50, 75, and 90th percentiles. The mean line is marked with a dotted line and all outliers are shown as dots.](image)

Figure 8-12: Distribution of sound class independent vote score for the first four voting comparisons. Box and whisker plots show the 10, 25, 50, 75, and 90th percentiles. The mean line is marked with a dotted line and all outliers are shown as dots.

8.4.2 Group voting preference

In order to summarise the voting data across the group of subjects, the median normalised vote score was calculated for each sound class within each comparison. The median was used in preference to the mean given the data did not commonly follow a normal distribution and the median was considered to be a better representation of the central tendency of the distribution in each sound class. In order to remove any bias due to unequal number of votes between subjects (and between sound classes), normalised vote scores (from each subject in each sound class) provided equal weighting in the calculation of the median score. That is, the number of votes was not used to weight the data points. In order to avoid potential bias in the median due to data points with relatively few votes, data points with two (or fewer) contributing votes were excluded from the calculation of the median. Note that these excluded data are shown in Figure 8-8 as small bubbles, but do not contribute to the summarised median data presented in Figure 8-11. Only the three dominant sound classes (Quiet, Speech in Noise, Noise) are shown in Figure 8-11 due to the small number of votes in the other classes.

The median scores are plotted with each voting comparison on a separate panel. The x-axis uses the same scale as the bubble plots and marked with stars to indicate statistical significance according to a one-sample singed rank test with hypothesised median score of zero. A table accompanies each panel describing the number of individual subjects with sufficient votes to be included in the analysis. Subject numbers are divided into categories based on the normalised vote score as ‘For’ or ‘Against’ the experimental algorithm (SpatialNR or SNR-NR). Divisions were made such that a vote score greater than zero was categorised as ‘For’, less than zero was categorised as ‘Against’, while a score equal to zero was categorised as ‘Equal’. The numbers of subjects within each category are shown as Against/Equal/For in the table accompanying each panel. In addition, the total number of votes obtained in each
scene was averaged across the subjects whose data contributed to each median vote score, and is shown in the table.

In order to test for differences in voting patterns between the three predominant sound classes, a one-way ANOVA was performed on the normalised vote scores with sound class (Quiet, Speech in Noise, Noise) as the main factor. A separate analysis was performed for each voting comparison. In situations where the sound class was not found to influence the voting patterns, a further analysis was performed by combining data from all sound classes to calculate a sound class independent vote score. This was done for each subject by summing votes across all sound classes. The sound class independent vote scores for the group of subjects are summarised in Figure 8-12 as box and whisker plots. All subjects’ votes were included in the analysis since all subjects voted at least three times in total in each comparison. In voting comparisons where there was no significant effect of sound class on the voting pattern, the sound class independent vote scores were statistically analysed. This was done with a one-sample t-test comparing the sound class independent vote scores to a hypothesised mean of zero.

8.4.3 Questionnaire

The number of returned questionnaires was 14 of 15 for the Zoom-SpatialNR(MA14) comparison, 15 of 15 for the SCAN-SpatialNR(MA14) comparison, 5 of 6 for the SCAN-SpatialNR(MA7) comparison, 14 of 15 for the SCAN-NR comparison. A questionnaire was not administered for the Directionality comparison.

SSQ results

Separate paired t-tests (and signed rank tests) were performed on the ratings obtained for each of the 14 SSQ questions in each of the four comparisons. No significant difference in rating between programs was found for any question in any comparison. Further analysis was performed by averaging the data across questions in each pragmatic subscale\(^{16}\), and for the entire questionnaire by averaging across all 14 questions. There was no significant difference in rating between programs for any of the pragmatic subscales or for the entire questionnaire for any of the comparisons that were made.

Overall preference results

Individual subject overall preference scores in quiet and noisy conditions are shown for each of the four comparisons in Figure 8-13 and the group data is summarised with box and whisker plots in Figure 8-14. Negative values indicate a preference for the baseline condition in the comparison (Zoom or SCAN), whilst positive values indicate a preference for the experimental condition (SpatialNR or SNR-NR). Degree of preference is given by the magnitude of the score: \( |3| = “Much prefer”, |2| = “Moderately prefer”, |1| = “Slightly prefer”, 0 = “No difference”.

\(^{16}\)Speech in quiet (items 2 & 3), Speech in noise (items 1, 4, 5 & 6), Speech in speech contexts (items 7, 8, 9 & 11) and Multiple stream-speech processing and switching (10, 12 & 14).
Figure 8-13: Questionnaire-based overall preference scores in quiet and noisy conditions for (A) Zoom-SpatialNR(MA14), (B) SCAN-SpatialNR(MA14), (C) SCAN-SpatialNR(MA7) and (D) SCAN-NR comparisons. Negative score indicates preference for baseline condition and positive score indicates preference for the experimental condition. Degree of preference is given by the magnitude of the score: $|3| =$ "Much prefer", $|2| =$ "Moderately prefer", $|1| =$ "Slightly prefer", $0 =$ "No difference".
A one-sample signed rank test on the hypothesised group median score of zero “No difference” was done separately for speech and noise in each of the four comparisons, forming eight tests in total. The analysis revealed that out of the eight separate tests, seven of them revealed there was no significant preference, and one, SCAN vs. SpatialNR(MA14) in Noise, revealed a median score of +1 “Slightly prefer SpatialNR(MA14)” that was significantly different from zero (P=0.025).

8.5 Discussion

8.5.1 Overall laboratory-based performance

The statistical analyses support the overall contention that SpatialNR provided superior noise reduction performance in terms of speech intelligibility (SRT), sound quality (MUSHRA) and acceptable noise level (ANL) compared to the other directionality settings that were evaluated (Standard, Zoom, Beam(Z)). The spatial configuration of the competing noise had an influence on the SRT outcome of the evaluation, with SpatialNR providing most benefit in the rear-half noises, suggesting that testing in a range of different noise types may be helpful in predicting outcomes for CI recipients. These results suggest that SpatialNR is most useful at combating background noise, when the competing sources are to the sides and/or rear of the listener.
Sources that are close to the target direction of zero degrees are not well attenuated by the algorithm due to the spatial basis upon which noise reduction is based.

8.5.2 Effect of MA

The laboratory-based SRT results show that the MA parameter had an impact on speech intelligibility for the noise types in which SpatialNR was effective (the rear half noises) and can be used to graduate SpatialNR performance between maximum benefit (MA96) and zero benefit (MA0). Although the maximum speech intelligibility benefit was obtained at MA96 in the two rear half noise types, there was no significant difference between MA14 and MA96, suggesting that speech intelligibility performance may have plateaued by the time MA had reached -14 dB. The MUSHRA and ANL results convey a similar message with MA96 providing the best performance overall, although not significantly different from MA14 or MA7. This thorough understanding of the impact of the MA parameter on speech intelligibility, sound quality and acceptable noise level within the confines of the laboratory can be used in conjunction with field-based responses to determine the most effective and acceptable MA under various listening conditions.

8.5.3 Effect of SNR-NR

The SRT benefit of NR was not robustly demonstrated in combination with SpatialNR. The SRT benefit in 2-SWN-rear did not quite reach significance, suggesting that the NR algorithm may be less beneficial in this configuration compared to the benefit when used in combination with other directionality settings. The benefit of NR in combination with Standard, Zoom and Beam was previously shown to be 1.2 – 1.4 dB SRT in the same noise type (Hersbach et al., 2012). A direct comparison within the same group of subjects is probably needed to confirm if this difference exists. The lack of any benefit in modulated noise types (4-TB-rear, 4-TB-full, 20-TB-full) is expected and previously demonstrated for the NR algorithm (Hersbach et al., 2012). Importantly, there was no measurable detrimental effect in any noise type of combining SpatialNR and NR in this study. In fact, ANLs were significantly improved by the NR algorithm, supporting the simultaneous use of the two algorithms for the purpose of combating both spatially separated and steady-state noise.

8.5.4 Effect of acclimatisation

There was no significant effect of test session on SRTs for Zoom or SpatialNR, nor on the difference in SRTs between Zoom and SpatialNR over the period of the experiment. This suggests that any acclimatisation to SpatialNR played only a minor role and did not impact speech intelligibility, either positively or negatively. That is, there was no measurable effect of exposure to the SpatialNR algorithm over the course of the experiment.
8.5.5 Effect of implementation platform

There was no significant difference in SRTs between any pair of conditions that used the two different platforms to implement the same processing condition. These results give confidence that any differences that exist between the implementation on the xPC platform and the implementation on the BTE platform are minor if any, and do not lead to measurable performance differences. Therefore, for practical purposes, the implementation platform was not considered to have an impact on the outcome of evaluation.

8.5.6 Normal hearing benchmark

A comparison between the NH group and the CI group using SpatialNR(MA14) processing revealed that, while SpatialNR processing generally improved the CI group performance, the performance did not match that of NH listeners in the same conditions. This indicates that the SpatialNR processing did not restore speech intelligibility to NH ability.

The two groups were differently affected by the different noise types. Specifically, the difference between the two groups was greater in 4-TB-rear and 4-TB-full compared to 20-TB-full. This effect is likely associated with ability to ‘listen in the gaps’, which is known to differ dramatically between CI and NH listeners. NH listeners are able to take advantage of temporal gaps in the fluctuating masker, while CI users’ ability to temporally resolve modulated competing noise is deprived due to their limited spectral resolution and reliance on envelope-based processing strategies (Nelson et al., 2003; Qin and Oxenham, 2003).

The difference between 4-TB-rear and 4-TB-full was not significant for NH group or CI group indicating that neither group found advantage in the alternate spatial separation of four competing talkers.

8.5.7 Voting preference

The bubble plots (Figure 8-8 and Figure 8-9) provide visualisation of individual subject voting preference, allowing comparison across different voting comparisons. An interesting example was S2 who showed a strong preference for SpatialNR in both voting comparisons that involved SpatialNR, a strong preference for SNR-NR and a strong preference for Beam in the directionality comparison. This subject presumably found benefit associated with all noise reduction processing generally preferring it to an absence of such processing. A contrary example was subject S23, who did not show an overall preference for SpatialNR in either comparison, nor did they show a preference for SNR-NR when that comparison was made. This suggests that S23 did not like any form of noise reduction, seemingly preferring an unaltered sound signal.

Subjects appeared to vote reasonably consistently across the different sound classes (Figure 8-10). This was indicated by the significant model fit to the voting scores in comparisons between most sound classes. One exception was the Quiet sound class in the Zoom-SpatialNR(MA14) comparison which was not significantly associated with either the Speech in Noise or Noise class. Aside from that exception, there was significant association between the
voting patterns across the three sound classes, Quiet, Speech in Noise and Noise. This indicates that subjects voted similarly in different situations, suggesting their choice of listening program was not significantly related to the acoustic environment as determined by the automatic classifier, SCAN.

While individual results provide some insight, the remainder of the discussion focusses on the group results for each voting comparison.

**Zoom-SpatialNR(MA14)**

The first comparison that was conducted for all subjects was Zoom vs. SpatialNR(MA14). The median normalised vote scores did not provide any significant difference between the two programs in any of the sound classes, although the number of subjects categorised as For/Against was 10/4 in the Noise class, showing some subjects preferred SpatialNR.

The one-way ANOVA revealed no significant effect of sound class (P=0.111) suggesting that the voting patterns were reasonably consistent across sound class for this comparison. This was contradicted somewhat by the analysis of individual data (Figure 8-10) that showed a lack of significant correlation between the Quiet class and the other two sound classes (Speech in Noise and Noise). Nevertheless, the sound class independent vote scores were calculated and the group mean was 0.20. The one-sample t-test revealed no significant difference from a mean of zero (P=0.204), indicating there was no difference between the two programs when summed across all sound classes.

Since the Zoom-SpatialNR(MA14) comparison was the first time subjects had performed the voting task over a period of take-home use, there was likely some learning involved in the task. This was indicated via verbal feedback from subjects, leading to a treatment of the data from this comparison as practice data. This in conjunction with the fact that SCAN was not fully operational in the comparison led to a lowering of the importance of this comparison’s contribution to the overall findings.

**SCAN-SpatialNR(MA14)**

The second comparison that was conducted for all subjects was SCAN vs. SpatialNR(MA14). This comparison was closer to the default operation of CP900, in which SCAN changed the underlying directionality depending on sound class, representing more closely the benefits of SpatialNR in a fully automated processor. Two sound classes (Speech in Noise and Noise) provided median normalised vote scores that were significantly different from zero. The median was 0.20 in Speech in Noise (P=0.037) and 0.42 in Noise (P=0.021). In the Quiet class, the median vote score was 0.64, but the result was not significantly different from zero (P=0.426) due to the spread of individual data. In all sound classes, the number of subjects voting For/Against SpatialNR(MA14) was 6/3 in Quiet, 9/1 in Speech in Noise and 10/2 in Noise.

Even though there were differences in voting patterns of the group between the different sound classes, the one-way ANOVA with sound class as the main factor revealed no significant
effect of sound class ($P=0.957$). This suggests that the voting pattern for the group of subjects was reasonably consistent across sound class for this comparison. This was supported by the individual subject analysis that showed vote scores in the three sound classes were significantly associated with one other (Figure 8-10).

The sound class independent vote scores were calculated and the group mean was 0.25. The one-sample $t$-test revealed the mean was not significantly different from a mean of zero ($P=0.088$) indicating there was no difference between the two programs when votes were summed across all sound classes. This is likely due to the voting patterns of two subjects (S17 and S25) in the Quiet class who did not prefer SpatialNR.

The data from this comparison indicate an overall group preference for SpatialNR in noisy conditions, with a majority of subjects preferring SpatialNR. Nevertheless, there were some individual subjects with preference against SpatialNR.

**SCAN-SpatialNR(MA7)**

There were six (of 15) subjects who reported an overall preference that was considered non-favourable for SpatialNR(MA14) and only those subjects took part in the evaluation comparing SCAN vs. SpatialNR(MA7). A low number of votes were obtained in the comparison and a median vote score could only be calculated for the Quiet sound class, in which the median score was zero, and the number of subjects For/Against SpatialNR(MA7) was 2/1.

The one-way ANOVA with sound class as the main factor revealed there was not sufficient data to run the analysis due to the low number of votes in some sound classes. Nevertheless, there were sufficient votes to calculate sound class independent vote scores and the group mean was 0.18. The one-sample $t$-test revealed the mean was not significantly different from a mean of zero ($P=0.307$) indicating there was no difference between the two programs when summed across all sound classes.

The results from this comparison suggest the less aggressive MA setting of $-7$ dB was acceptable for this group of subjects, probably because the algorithm provided little noticeable effect or noise reduction.

**SCAN-NR**

The SCAN-NR comparison did not involve SpatialNR, but was designed to evaluate the SNR-NR algorithm instead. In this way it provided voting information on a more well-known algorithm, which could then be used to illustrate the degree of preference obtained with other comparisons. The median vote score was 0.75 in Quiet ($P=0.129$), 0.50 in Speech in Noise ($P<0.001$) and 0.71 in Noise ($P=0.005$) while the number of subjects For/Against SNR-NR was 7/2 in Quiet, 13/1 in Speech in Noise and 10/2 in Noise.

Although there appeared to be differences in voting patterns between the different sound classes, the one-way ANOVA revealed no significant effect of sound class ($P=0.355$) suggesting that the voting patterns were reasonably consistent across sound class for this comparison.
This was supported by the individual subject analysis that showed vote scores in the three sound classes were significantly associated with one other (Figure 8-10). The sound class independent vote scores were calculated and the group mean was 0.42 which was significantly different from zero according to the one-sample t-test (P=0.004). This indicates there was a significant difference between the two programs when summed across all sound classes, with subjects preferring the SCAN-NR program over the SCAN alone program.

These results represent a relatively strong preference for SNR-NR, at least in the Speech in Noise and Noise classes, confirming that the reported preference for SNR-NR is indeed reflected in the voting data that was collected.

**Directionality**

The Directionality comparison was performed using three programs, so the analysis was slightly different to the previous comparisons. In order to summarise the group’s data, the median was calculated directly from the proportion of votes for each of the three programs, Standard, Zoom and Beam(Z). In the Quiet class, the median proportion of votes was 0.17 for Standard, 0.28 for Zoom, and 0.48 for Beam(Z). Wilcoxon signed rank tests were used to compare each pair of processing conditions, Standard/Zoom, Standard/Beam(Z) and Zoom/Beam(Z). A significant difference between Standard and Beam(Z) was found (P=0.048) and the differences were not significant for Standard/Zoom (P=0.191) and Zoom/Beam(Z) (P=0.787). It is interesting that votes were predominantly recorded only in the Quiet class for the Directionality comparison, with relatively very few votes recorded in the noisy sound classes, even though the time spent by users in these classes was approximately equal. This is markedly different to the pattern of voting obtained in all previous comparisons, where the number of votes in each of the three sound classes was approximately in proportion to the time spent in each sound class. The reason for this may be associated with the noticeable effect of directionality, which appears to be more salient in the Quiet class than in the noisy classes.

Due to the lack of votes outside the Quiet class, further analysis upon the sound class independent vote scores was not conducted for the Directionality comparison.

### 8.5.8 Voting patterns

There was some variation in the number of votes recorded over each comparison period, and the median number of votes for an individual when summed across all sound classes was between 9 and 15 votes, depending on the comparison. The distribution of votes across listening scenes was somewhat related to the time spent in each scene (Figure 8-15), although the pattern varied depending on the comparison type. Even though the number of votes in each sound class changed with the comparison type, the time spent in each class was relatively consistent across comparisons.
Figure 8-15: Number of vote events vs. number of listening hours, separated by take-home comparison on separate panels. The sound class is identified by the marker colour, blue=Quiet, green=Speech in noise and orange=Noise. Markers show the median and error bars show the first and third quartiles.

The median proportions were Quiet 50-54%, Speech 8-10%, Speech in Noise 20-25%, Noise 13-16% and Music 1-3%. The number of votes in each sound class did not always increase with the time spent in each class; that is, the voting rate was not consistent across sound class and varied with the comparison being made.
Figure 8-16: Rate of voting (votes/hr) compared with rate of changing programs (changes/hr) for each of the voting comparisons on separate panels. The sound class is identified by the marker colour, blue=Quiet, green=Speech in noise and orange=Noise. Markers show the median and error bars show the first and third quartiles.

In the Directionality comparison, very few votes were recorded outside the Quiet sound class. Within the Quiet class, more votes were recorded than in any of the other comparisons. This was partly associated with the time spent in the Quiet class during that comparison, which was more than during the other comparisons. Accounting for the time spent by calculating the voting rate (votes/hr), it can be seen that the voting rate in the Quiet class in the Directionality comparison was similar to the voting rate in the noisy classes during other comparisons (Figure 8-16). The reason for a lack of votes in the noisy classes (Speech in Noise and Noise) in the Directionality comparison was possibly due to lack of perceived difference between programs.
under noisy conditions. This fits with the instruction given to subjects to not vote if no program was superior to the others. This is also supported by the number of program changes that were made (Figure 8-16) where the low number of votes in the noisy classes during the Directionality comparison was not associated with low numbers of program changes. That is, it is unlikely the lack of votes were associated with a lack of trying the different programs. It appears a strong and perceptible difference was noted in the Quiet class, indicated by the high rate of voting. The underlying reason why changes in directionality led to perceptible differences in the Quiet class but not in the noisy classes seems unexpected, but was perhaps associated with the level of background noise and, in particular, the change to the background noise that resulted from the processing. It is likely that in the Quiet class, background noise was not completely absent but existed at a low level, such that changes in directionality were detectable because the background noise level changed, perhaps from a low level to a level that was inaudible. In noisy environments, changes to background noise level due to directionality processing may have been more subtle and not as readily apparent since the background noise would have been at a higher level in that sound class. For example, an attenuation of the background noise from 70 dB SPL to 65 dB SPL may not have been as detectable as a 5 dB change that altered the level from low to inaudible.

The lack of voting data in noisy situations for different directionality settings contrasts with the laboratory-based results, where Zoom and Beam(Z) provided significant improvements in speech intelligibility and sound quality compared to Standard. Given the positive outcomes in two of the three laboratory-based tests\(^{17}\), one might hypothesise that CI users might also report a preference for these algorithms in the field. The voting data revealed that CI users did prefer Zoom/Beam(Z) in the Quiet sound class, but no preference was found in the noisy classes, which does not fit with the hypothesis. The lab-based performance does not appear to predict the field-based voting preference.

Related conclusions were drawn in a study with hearing aid users wearing directional microphones (Walden et al., 2003; Walden et al., 2004). Amongst the conclusions was that “Directional microphones work better in the test booth than they do in everyday listening situations” (Walden et al., 2003), a contention supported by the current study in which laboratory-based directional benefit was not coupled with field-based preference in noise. However, Walden et al. (2004) found that omni-directional mode tended to be preferred in relatively quiet situations, and that directional mode tended to be preferred when background noise was present, with the signal source nearby and in front of the listener. Both of these results are contradicted in the current study in which directionality (Zoom/Beam) was preferred in quiet situations and no directional preference was found in noisy situations. Perhaps aspects of sound quality and spatial awareness, perceived differently by hearing aid and CI users, potentially influence directionality preference.

A contrast to the Directionality comparison is found in the SNR-NR data, where two of three laboratory-based metrics revealed no significant improvement\(^{18}\), but the field-based data revealed an overwhelmingly strong preference for SNR-NR according to the voting data. As

\(^{17}\) Improvements in ANL were not statistically significant

\(^{18}\) No significant improvement in speech intelligibility or sound quality in any noise type and a significant improvement in acceptable noise level only in the 2-SWN-rear noise type.
with the Directionality comparison, the lab-based and field-based data do not agree. However, it was the field-based preference that showed the benefit rather than the laboratory-based data. This pattern of results is also observed for digital noise reduction in hearing aids where the benefits are attributed to improved ‘ease of listening’ (Bentler, 2005).

The correspondence between SpatialNR field and laboratory-based data appears more conventional, where improvement in laboratory-based speech intelligibility, sound quality and acceptable noise level were accompanied by field-based preference for the algorithm. This pattern was not observed for the Directionality or SNR-NR comparisons.

8.5.9 Program listening time and voting preference

A further analysis was made on the relationship between voting patterns and listening time to determine if users spent more time listening to the program they preferred; i.e. was listening time related to voting preference. To do the analysis, the time spent in each program was converted to a proportion of the total listening time in each acoustic sound class. Furthermore, the proportion was scaled between -1 and +1, such that -1 indicated 100% of the time listening to the baseline program, +1 indicated 100% of the time listening to the experimental program and 0 indicated 50/50 allocation of listening time to the two programs. The normalised time proportion was therefore analogous to the normalised vote score and these are presented graphically for the Zoom-SpatialNR(MA14), SCAN-SpatialNR(MA14) and SCAN-NR voting comparisons in Figure 8.17.

The bubble size indicates the number of votes. Data points with less than three votes were removed from the analysis due to the questionable reliability of the vote score based on so few votes. A model of the form $y = ax$ was fit using least squares regression. The model parameter, $a$, was significant for each sound class in each voting comparison with one exception. The exception was the Quiet sound class in the Zoom-SpatialNR(MA14) comparison in which the model parameter was not significant. The results indicate that there was significant association between voting preference and listening time suggesting that subjects typically continued to use the program they had voted for. In the absence of a voting button to motivate the user to change programs, the relationship between program preference and listening time not hold.
Figure 8-17: Relationship between voting preference and listening time for the Zoom-SpatialNR(MA14), SCAN-SpatialNR(MA14) and SCAN-NR comparisons. The normalised proportion of listening time is plotted against the normalised vote score for each sound class. Bubble size indicates number of votes for each data point (<3 votes not shown). Colour indicates sound class (blue=Quiet, green=Speech in Noise, orange=Noise). Solid lines represent the model \( y=ax \). Dotted lines show the 95% C.I. for the estimate of the model parameter, \( a \). A positive gradient on both the upper and lower 95% C.I. indicates the model fit was significant.
8.5.10 Laboratory-based performance and voting preference

In this section, the relationship between laboratory and field-based assessment is analysed. The purpose is to establish whether the performance of individual subjects on laboratory-based tests was related to the field-based voting preference of those individuals.

The laboratory-based tests were compared to the field-based voting scores for each subject to assess whether any relationship could be established. To do this, the normalised voting scores for the comparisons Zoom-SpatialNR(MA14), SCAN-SpatialNR(MA14) and SCAN-NR were analysed. The voting scores were plotted against each of the laboratory-based measures, SRT, MUSHRA, and ANL (Figure 8-18 - Figure 8-22). The voting scores in these comparisons always pertained to a comparison between two listening programs so in order to compare those against the laboratory-based measures, the benefit scores were used. That is, the difference in performance between two conditions was used rather than the absolute score. For example, in the Zoom-SpatialNR(MA14) comparison, the two programs under comparison were Zoom and SpatialNR(MA14). The SRT benefit was calculated as the difference between the SRTs in these two conditions, which was then plotted against the voting score. The difference scores were also calculated for MUSHRA ratings and ANLs in a similar fashion.
Figure 8-19: Relationship between laboratory and field-based assessment for the Quiet sound class of the SCAN-SpatialNR(MA14) comparison. SRT, MUSHRA and ANL benefit was calculated as the difference between the SpatialNR(MA14) and Standard conditions. Normalised vote scores were from the Quiet sound class. Solid lines represent the least squares regression to the model $y = ax$. Dotted lines show the 95% C.I. for the estimate of the model parameter, $a$. A positive gradient on both the upper and lower 95% C.I. indicates the model fit was significant.

For the Zoom-SpatialNR(MA14) and SCAN-NR comparisons, the sound class independent vote score was used (Figure 8-18, Figure 8-22). That is, the vote score summed across all sound classes was used. For the SCAN-SpatialNR(MA14) comparison, the underlying directionality changed between sound classes during the field-based assessment. Therefore, the vote scores from each sound class were plotted against the laboratory-based benefit pertaining to the relevant comparison that was made. Specifically, in the Quiet sound class the benefit over Standard was used (Figure 8-19), in the Noise class the benefit over Zoom was used (Figure 8-20), and in the Speech in Noise class the benefit over Beam(Z) was used (Figure 8-21).

In cases where the laboratory-based outcome was significantly affected by the noise type used for evaluation, each of the noise types (2-SWN-rear, 4-TB-rear, 4-TB-full and 20-TB-full) were plotted separately. Otherwise, the outcome measure was averaged over all the noise types. For the Zoom-SpatialNR(MA14) and SCAN-SpatialNR(MA14) comparisons, the SRTs were plotted for each noise type. For the SCAN-NR comparison, SRTs were averaged over all noise types. MUSHRA ratings and ANLs were averaged over all noise types for all comparisons.
In order to test for a statistically significant relationship between a laboratory-based outcome measure and the voting score, a linear model of the form \( y = \alpha x \) was fitted using least squares regression (solid line). The 95% C.I. (broken line) was used to determine a significant fit of the model to the data. Therefore, when the upper and lower C.I. had positive gradients, the parameter, \( \alpha \), was significantly different from zero, and therefore a significant relationship was demonstrated. In all cases when the 95% C.I had a negative gradient, the estimate of the model parameter, \( \alpha \), was not significantly different from zero, and therefore no significant relationship was established.

Across the three comparisons, only three instances were identified in which the laboratory-based outcome measure provided a significant prediction of the field-based voting preference. That is, a significant relationship was only established in a few specific situations. Those significant relationships were between voting preference and:

- SRT benefit (4-TB-rear) of SpatialNR(MA14) re. Zoom (Figure 8-20),
- SRT benefit (4-TB-rear) of SpatialNR(MA14) re. Beam(Z) (Figure 8-21) and
- ANL benefit (all noises) of NR (Figure 8-22).
Figure 8-21: Relationship between laboratory and field-based assessment for the Speech in Noise sound class of the SCAN-SpatialNR(MA14) comparison. SRT, MUSHRA and ANL benefit was calculated as the difference between the SpatialNR(MA14) and Beam(Z) conditions. Normalised vote scores were from the Speech in Noise sound class. Solid lines represent the least squares regression to the model $y = ax$. Dotted lines show the 95% C.I. for the estimate of the model parameter, $a$. A positive gradient on both the upper and lower 95% C.I. indicates the model fit was significant.

Figure 8-22: Relationship between laboratory and field-based assessment for the SCAN-NR comparison. SRT, MUSHRA and ANL benefit was calculated as the difference between the SpatialNR(MA14)NR and SpatialNR(MA14) conditions. Normalised vote scores were summed across all sound classes. Solid lines represent the least squares regression to the model $y = ax$. Dotted lines show the 95% C.I. for the estimate of the model parameter, $a$. A positive gradient on both the upper and lower 95% C.I. indicates the model fit was significant.
8.5 Discussion

It is worth noting that MUSHRA sound quality ratings were close to significant in predicting the voting preference in the noisy classes (Speech in Noise and Noise) during the SCAN-SpatialNR(MA14) comparison (Figure 8.18 and Figure 8.19). Although the relationship was not significant, there may be some potential for MUSHRA sound quality ratings to provide some indication of voting preference in the field for individual subjects.

The analysis of individual results indicates that the relationship between laboratory and field-based metrics was not strong at the individual subject level. It is unlikely that users made their voting decisions based entirely on speech intelligibility, sound quality or acceptable noise level alone. Donaldson et al. (2009) suggested that speech intelligibility and ANL were not systematically related to one another19, but taken together accounted for 72% of the variance in responses to global APHAB scores, an indicator of overall preference. However, in this study, the field-based voting exercise elicited responses that were not necessarily evident in the laboratory. This could be due to limitations of reproducing real-world environments within the laboratory, a function of the differences between the psycho-physical tasks and real-world preference, or a combination of both. Given that the group-based analysis established significant laboratory-based performance benefit and significant field-based preference outcomes, it is possible that the lack of predictive relationship at the individual subject level was due, at least in part, to the inherent variability of the laboratory-based metrics.

8.5.11 Questionnaire

As opposed to the outcomes found with the voting data where there was a clear preference for SpatialNR and SNR-NR, the SSQ ratings indicate the questionnaire was not sensitive to the differences between algorithms that were evident in the voting data. Perhaps this was partially due to the general nature of questionnaire-based assessment, which required the subjects to recall particular listening situations and their program preference therein. The additional questions posed on overall preference in quiet and noisy conditions yielded more conclusive outcomes than the SSQ component and were compared against the voting data.

In order to compare the questionnaire-based overall preference data with the voting data the individual subject overall preference score and voting score were plotted against one another (Figure 8.23) for the Zoom-SpatialNR(MA14), SCAN vs SpatialNR(MA14) and SCAN-NR comparisons. Overall preference score in quiet was plotted against the normalised voting score for the quiet class and the overall preference score in noise was plotted against the normalised voting score for the noise class. The number of votes is indicated by the size of the marker bubble and data points with less than three votes being removed due to questionable reliability of the vote score associated with such few votes. A model of the form $y = ax$ was fitted using least squares regression for the Quiet and Noise data separately for each comparison. A significant association was found in Quiet for the Zoom-SpatialNR(MA14) and SCAN-SpatialNR(MA14) comparisons, but not the SCAN-NR comparison. In Noise, there was a significant association in the Zoom-SpatialNR and SCAN-NR comparison, but not in the SCAN-SpatialNR(MA14) comparison.

---

19 in a group of 20 CI users
Figure 8-23: Correlation of overall preference score and normalised vote score for the Zoom-SpatialNR(MA14), SCAN-SpatialNR(MA14) and SCAN-NR comparisons. Both axes indicate degree of preference where negative scores indicate a preference for SCAN and positive scores a preference for SpatialNR(MA14) or NR. Size of bubble indicates number of votes recorded for that data point (<3 votes not shown). Colour of bubble indicates the listening scene (blue=Quiet, orange=Noise). Solid lines represent the model $y=ax$. Dotted lines show the 95% C.I. for the estimate of the model parameter, $a$. A positive gradient on both the upper and lower 95% C.I. indicates the model fit was significant.

The lack of a consistent relationship between the questionnaire-based preference and the voting-based preference indicates there were differences in the way some subjects responded to the two tasks. This could be associated with the time lapse between exposure and when the preference was made, either instant on-the-spot in the case of voting, or delayed until visit to the laboratory in the case of questionnaire. Additionally, the questionnaire preference was made at the conclusion of a two week listening period, and the subject was asked for a single
response, and so represents the subjects’ perceived average over the period. The voting score on the other hand, was generated from many individual voting responses, and whilst the data was averaged over the same period, it did not rely on the subject’s recollection to perform the averaging. It is also possible that when answering the questionnaire, subjects confused their preference of P1 and P2, since they were required to recall their program preferences over a two week period. The voting task provided a more direct selection of program choice in this respect.

8.6 Conclusions

The SpatialNR algorithm was subjected to a host of laboratory and field-based assessments. These assessments provide complementary for determining the effectiveness of the algorithm in mitigating background and competing noise. Overwhelmingly, the data indicated the superior performance of SpatialNR compared to the other directionality modes that were evaluated. In particular, the laboratory-based tests of speech intelligibility, sound quality and acceptable noise level all demonstrated that SpatialNR provided superior performance. This was particularly evident when the competing noise was to the sides and rear of the listener. The benefit was immediate, in that there was no time needed for CI users to become accustomed to the change in processing. Also, their speech intelligibility performance did not change over the course of the experiment, nor with the platform used to implement the algorithms. The complementary field-based preference data also indicated the most subjects preferred the program with SpatialNR, rather than without, although some subjects did not prefer the SpatialNR program. Those subjects who did not were further evaluated with an adjustment that provided less aggressive noise reduction from the SpatialNR algorithm, and the results showed that only one subject continued to vote against SpatialNR. The laboratory-based results with different MA settings suggest that the less aggressive setting (MA7) did not provide as much noise reduction benefit as MA14. However, the field-based assessment showed that this setting was acceptable to 5 of 6 subjects who did not prefer the more aggressive setting. It seems feasible that the MA setting could be adjusted for individual subjects, according to their preference, notwithstanding the fact that there is a trade-off with performance as the MA setting approaches zero.

While preference for SpatialNR differed among individuals, it was hypothesised that preference might vary with the acoustic sound class. This was not well supported in the field-based preference data. The comparisons made with SpatialNR did not reveal voting patterns that could be easily separated by sound class – individual variability contributed more than differences between sound classes.

In the laboratory, SNR-NR was evaluated in conjunction with SpatialNR in order to demonstrate how a product with both features might perform. Although the data supported a benefit of using the two algorithms together, the additional benefit of SNR-NR was only detected in the acceptable noise level test. During the field-based assessment, SNR-NR was evaluated in combination with Standard, Zoom and Beam(Z), depending on the sound class, but was not directly evaluated in combination with SpatialNR in the field. The field-based
preference data was overwhelmingly positive, reflecting the fact that voting decisions were made on aspects of listening that were not well re-produced in the laboratory-based tests. The relationship between laboratory and field-based outcomes was further complicated during the field-based directionality comparison in which almost no votes were recorded in noisy situations. In contrast, the laboratory-based evaluation showed a clear benefit of the directional algorithms in noisy conditions. Possible reasons for the disconnect between laboratory and field-based outcomes could be associated with differences in the acoustic environment used for evaluation as well different psychophysical aspects of listening in noise involved in the different tasks.

The system of voting used in this project allowed users to provide direct feedback about their preferred listening program, something which was not possible with the questionnaire. Clear, statistically significant preferences were evident in the voting data where subjects made their program choice while they were actually listening. This real-time method of data collection concerning subjective preference appears to be a more direct assessment of individual choice as it did not require subjects to later recall specific acoustic situations and program preference, as demanded via the questionnaire.

The voting task likely motivated users to change programs and actively make a decision about preference. Although not directly addressed in this experiment, the act of voting was thought to be significant in motivating subjects to try the different listening programs in the field. This may not have been the case without the vote button, which provided a psychological end point of the task, allowing users to confirm their choice after exploring different programs. Without the vote button, the task of trialling programs would have been more open-ended, without any means for confirming choice aside from retaining the selection of the preferred program. With no vote button, the task of determining listening preference would likely rely assuming that program usage time was associated with preference, which may not always be valid. Problems exist with this interpretation in situations where there was no underlying program preference (e.g. Directionality comparison in noisy conditions), where listening time was associated with other factors and not related to indication of preference. In addition, it would have been difficult to account for periods during the day where a user was not actively evaluating the different programs. The vote button provided a way to mark precise time points during the experiment that were associated with direct evaluation of different programs. Without a vote button, there was no way to distinguish between periods of direct evaluation and periods of non-evaluation where a user may have left the device in one program for a period of time. The inclusion of the vote button likely provided a more direct indication of program preference than could have been achieved simply by providing different listening programs without a means to vote.
Chapter 9

Summary and conclusions

During this research project, a noise reduction algorithm for cochlear implants (CIs) has been developed and evaluated. The design of the algorithm encompassed aspects from single-microphone and multi-microphone solutions, with the aim of creating a robust algorithm for use in CIs.

Unlike an adaptive beamformer, the SpatialNR algorithm did not work on a null steering basis. Instead, the SNR was estimated by analysing the ratio of front and rear facing directional microphones, and the SNR was used to apply a smooth attenuation function to remove the noise. Thus, the SNR estimate provided a method of selecting the most important frequency channels for speech understanding. The SpatialNR algorithm was expected to have distinct advantages over other multi-microphone approaches based on null-steering or filtering (e.g. beamforming and multi-channel Wiener filtering). It is well known that these methods are excellent at cancelling a single masker that is spatially separated from the target speech, particularly in low reverberation conditions. However, as the number of competing maskers increases and the masking sound becomes more diffuse, and with increase in reverberation, the effectiveness of these algorithms is reduced. In essence, these algorithms cannot cancel many sounds from many directions at any one instance, but can only adapt a filter to minimise the total noise energy in the output signal. In contrast, the SpatialNR algorithm applies a spectral gain modification based on an SNR estimate. This difference means that multiple noise sources can be removed at a particular time instant, effectively allowing removal of multiple noise sources concurrently.

The SpatialNR formulation had similarities to a coherence-based spatial filter (Yousefian and Loizou, 2013) and a phase-based spatial filter (Goldsworthy et al., 2014) that have both been evaluated in CIs. The three algorithms were similar in that each calculated a metric that was representative of the direction of arrival (DOA) that was used to spatially filter the signal. The underlying properties used in the calculations were coherence, phase and directional microphone power. As a consequence, the performance of the algorithms were likely to vary depending on the acoustic test conditions and the differing assumptions made for determining DOA. For example, the coherence-based filter provided benefit in low reverberation but was found to degrade in performance when reverberation was introduced (Yousefian and Loizou,
2013). A possible reason for this is the dependence of coherence-based methods on the acoustic environment, including reverberation, noise source location and microphone array orientation (Martin, 2001b). The evaluation of the phase-based spatial filter was performed in reverberant conditions, but two of the three noise sources were located directly in the spatial null of the fixed directional pattern (Goldsworthy et al., 2014), potentially limiting generalisation of the findings more broadly. The acoustic situation, spatial location of the noise sources (static or otherwise), level of reverberation and speech recognition task itself (sentences, words or phonemes) all contribute to the reported outcomes. This highlights the importance of evaluating across a range of conditions in order to provide an informative description of algorithm performance.

The evaluation of the SpatialNR algorithm encompassed five experiments (Chapters 4 to 8) that were primarily concerned with objectively quantifying speech understanding of CI users in a variety of controlled acoustic environments within the laboratory. However, in the final experiment (Chapter 8), subjects were able to make comparisons between different program settings ‘in the field’ by voting for their preferred setting in situ. That is, subjects had the ability to change programs and choose their preferred program whilst using their device away from the laboratory. The outcomes of the final experiment provide critical information concerning subjective preference of noise reduction settings which, when coupled with laboratory-based perception tests, gives a more thorough understanding of the algorithm.

The laboratory-based experiments presented in Chapters 4 to 7 concentrated on assessing the speech intelligibility benefit provided by the noise reduction algorithm. Evaluations took place in a variety of noisy configurations, in rooms with varying reverberant properties and under conditions of microphone mismatch. All of these evaluations were designed to provide real-world relevant situations in which CI users may benefit from noise reduction. The noise configurations used speech weighted noise or multiple competing talkers. The spatial configuration of the noise sources included competing noise from the rear hemi-field or from the full circle, providing differing examples of noisy situations that CI users may encounter in their daily lives. Given the known issues of multi-microphone noise reduction with reverberation and microphone mismatch, these two practical issues were evaluated in detail. The reverberant properties of the room and microphone mismatch were systematically varied in order to understand the impact on laboratory-based perceptual measures.

The SpatialNR algorithm demonstrated robust performance over a range of noisy environments and in reverberation. An unexpected finding was that the benefit was increased in reverberation. This was unexpected because other studies have documented a clear decrease in benefit in reverberation for directional processing. This finding suggests that the first line of defence in reverberant environments could be the use of directional processing to reduce noise. The robust performance benefit across a wide variety of situations suggested that CI users could benefit from the device. This was confirmed through the voting and questionnaire data obtained during the final take-home investigation.

Microphone mismatch was found to have a detrimental effect on the processing. The benefit from directional processing decreased due to microphone mismatch. This suggests that addressing the area of microphone matching would be advantageous in terms of the benefit
that directional algorithms can provide. The data provide evidence supporting a recommendation that 4 dB mismatch removes all directional benefit, and that 2 dB is tolerable as the maximum acceptable mismatch before significant directional degradation occurs.

While there was a focus in this study on evaluating across a wide range of acoustic situations, algorithm parameter variations were also assessed. These included the bias parameter, smoothing parameter and maximum attenuation parameter. The bias parameter was found to change performance due to the change in spatial filtering properties and relative proportions of type I and type II errors, with an optimal setting of 3 dB. The smoothing parameter was found to alter performance, mostly impacting sound quality, with asymmetrical smoothing of 5/50ms found to be acceptable for this application. The maximum attenuation parameter, used to smoothly turn the algorithm on and off, operated by mixing the noisy input signal and the noise reduced output signal in different ratios. Predictably, when there was more noisy signal in the mix, the benefit of the algorithm was reduced. Characterisation of laboratory-based perception at different maximum attenuation settings provides a thorough understanding of the benefit provided in controlled listening conditions.

The final experiment presented in Chapter 8 culminated in a BTE device implementation that facilitated field-based assessment of the algorithm. The findings indicate that preference for noise reduction processing varied across individuals, and was not systematically associated with the automatically classified acoustic scene. The outcome suggests that providing a means by which individuals can alter the strength of noise reduction may be beneficial. Such a system could exploit the voting data to choose the best parameters for the population, and ultimately allow individuals to choose their own settings.

9.1 Future research

The SpatialINR noise reduction algorithm works on the basis of creating an estimate of the SNR, which in this algorithm is fundamentally based on direction of arrival or, more specifically, the front-to-back ratio. An advantage of directly estimating the SNR means that integration with other algorithms that also estimate SNR may be possible, in a manner that adds robustness and accuracy to the overall SNR estimate. For example, a single microphone noise reduction algorithm that estimates the SNR could be used. It cannot estimate the SNR based on direction or spatial information since there is only access to a single microphone, but relies on the assumption that speech fluctuates in amplitude more quickly than noise. In some situations this assumption is violated, for example when the competing source is speech. Similarly, the assumption made by SpatialINR that the target is in front will also be violated on occasions. However, combining both spatial and statistical information to estimate the SNR could minimise occasions when such assumptions are violated. This would improve the overall performance of the system across a wider variety of signal inputs.

The SpatialINR algorithm has been developed specifically for use in CI sound processing, but may be transferrable to acoustic applications such as hearing aids. Two similar algorithms based on spatial filtering were evaluated in acoustic applications and were found to give superior performance compared to the main beamformer (Cao et al., 1996) and compared to
beamformer with Wiener post-filter (Wolff and Buck, 2008). This indicates the SpatialNR algorithm may be used in acoustic applications with a likelihood of success. However, the bias parameter used to tune the aggressiveness of noise reduction, which changes the relative amounts of speech distortion and noise reduction, would probably need to be tuned differently for acoustic applications. For example, Brungart et al. (2006) showed that normal hearing listeners perform best with a negative gain threshold, whilst CI listeners perform best with a positive gain threshold (Mauger et al., 2012b). This indicates that acoustic applications may need a less aggressive (lower) bias value, maintaining speech quality at the expense of noise reduction performance. Another issue that relates to acoustic processing, not investigated in the present study, is reconstruction of the time-domain audio signal through the use of an inverse FFT. CI processing requires no such transformation because the frequency channel energy is used to determine the level of electric stimulation delivered by the CI electrodes. Reconstruction of a time-domain signal could introduce unwanted artefacts if the noise reduction gains change too quickly over time or across neighbouring frequencies.

The voting paradigm used in Experiment V (Chapter 8) allowed users to change listening programs and vote for their preferred setting. The system was based on using the CP900 listening programs configured with parameter settings that varied from program to program. The program-based approach required the user to change programs which incurred a time delay between switching of approximately 1 s. Furthermore, the program-based system was limited to a maximum of four discrete programs that were available on the CP900 device. Therefore, for the purpose of exploring a continuously variable parameter like maximum attenuation, the program-based approach could be improved. An alternative solution could provide a means for directly controlling the parameter under investigation, for example by providing an interface similar to a volume control, which would allow the user to adjust the parameter value up and down in a continuous manner. In this way, a broader range than four discrete values could be evaluated, and the mechanism for changes would not involve the same time delay associated with program-based changes. The vote button would have the same function as in the program-based system, providing a means for users to confirm their most preferred setting.

9.2 Final summary

The main objectives of this research were

- to design a multi-microphone noise reduction algorithm to improve CI performance in noisy conditions,
- to evaluate the algorithm using speech tests with CI users, and
- to obtain listening preference for the algorithm from users in their day-to-day lives.

To achieve these objectives, the algorithm was described in detail from a theoretical perspective, which provided an insight into the operation of the algorithm and the influence of varying its parameters. Throughout the ensuing experiments with CI users, parameter variations were assessed from a perceptual point of view in order to define suitable settings. The maximum attenuation parameter ultimately provided the main control point for smoothly
engaging the noise reduction algorithm and remains the parameter that will likely be suited to individual user adjustment.

The algorithm was evaluated in the laboratory with CI users primarily through speech intelligibility tests, but also through sound quality and acceptable noise level tests. The algorithm was compared against existing algorithms from Cochlear Ltd and was shown to provide enhanced speech intelligibility, sound quality and allow a higher acceptable noise level. In particular, a significant advantage in speech intelligibility was demonstrated in situations when the competing noise was from the sides and rear of the listener. Algorithm performance was maintained and unexpectedly increased in reverberant environments. As with other multi-microphone algorithms that were evaluated, performance was negatively affected by microphone mismatch.

Ultimately, the algorithm was implemented on a BTE sound processor which allowed users to evaluate the algorithm away from the laboratory. A system was developed where users indicated their preference using their processors’ remote control, which showed that CI users predominantly preferred the noise reduction algorithm. This corroborates the expected advantage indicated by laboratory-based testing.

The laboratory-based metrics and field-based preferences tended to provide complimentary information about the ‘benefit’ of noise reduction. It appears that no single metric, in any single acoustic condition, was sufficient to describe the ‘benefit’ perceived by the cochlear implant user. Each of the different measures provided a different insight into the ‘benefit’ and all can be used to guide the final use of the algorithm. The performance benefit demonstrated in this work through laboratory and field-based evaluation suggests that CI users would be likely to obtain ‘benefit’ from incorporation of the algorithm in their CI sound processor.
Appendix – Questionnaire

Speech Spatial Qualities

Advice about answering the questions

The following questions inquire about aspects of your ability and experience hearing and listening in different situations.

For each question, put a mark, such as a cross (x), anywhere on the scale shown against each question that runs from 0 through to 10. Putting a mark at 10 means that you would be perfectly able to do or experience what is described in the question. Putting a mark at 0 means you would be quite unable to do or experience what is described.

As an example, question 1 asks about having a conversation with someone while the TV is on at the same time. If you are well able to do this then put a mark up toward the right-hand end of the scale. If you could follow about half the conversation in this situation put the mark around the mid-point, and so on.

We expect that all the questions are relevant to your everyday experience, but if a question describes a situation that does not apply to you, put a cross in the “not applicable” box. Please also write a note next to that question explaining why it does not apply in your case.

Please answer the following questions, then go on to the questions about your hearing

Your name:

Today’s date:

Your age:

Please check one of these options:

- I use one Sound Processor (left)
- I use one Sound Processor (right)
- I use two Sound Processors

If you have been using Sound Processor/s, for how long?

- _____ years
- _____ months
- _____ weeks

If you have two processors and have used them for different lengths of time, please write down both.
Speech Spatial Qualities (Part 1: Speech hearing)

1. You are talking with one other person and there is a TV on in the same room. Without turning the TV down, can you follow what the person you’re talking to says?

<p>| | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>Not at all</td>
<td>Perfectly</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>P2</td>
<td>Not at all</td>
<td>Perfectly</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
</tr>
</tbody>
</table>

2. You are talking with one other person in a quiet, carpeted lounge-room. Can you follow what the other person says?

<p>| | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>Not at all</td>
<td>Perfectly</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>P2</td>
<td>Not at all</td>
<td>Perfectly</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
</tr>
</tbody>
</table>

3. You are in a group of about five people, sitting round a table. It is an otherwise quiet place. You can see everyone else in the group. Can you follow the conversation?

<p>| | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>Not at all</td>
<td>Perfectly</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>P2</td>
<td>Not at all</td>
<td>Perfectly</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
</tr>
</tbody>
</table>

4. You are in a group of about five people in a busy restaurant. You can see everyone else in the group. Can you follow the conversation?

<p>| | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>Not at all</td>
<td>Perfectly</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>P2</td>
<td>Not at all</td>
<td>Perfectly</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
</tr>
</tbody>
</table>
Speech Spatial Qualities (Part 1: Speech hearing)

5. You are talking with one other person. There is continuous background noise, such as a fan or running water. Can you follow what the person says?

<table>
<thead>
<tr>
<th></th>
<th>Not at all</th>
<th>Perfectly</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>0 1 2 3 4 5 6 7 8 9 10</td>
<td>Not applicable</td>
</tr>
<tr>
<td>P2</td>
<td>0 1 2 3 4 5 6 7 8 9 10</td>
<td>Not applicable</td>
</tr>
</tbody>
</table>

6. You are in a group of about five people in a busy restaurant. You CANNOT see everyone else in the group. Can you follow the conversation?

<table>
<thead>
<tr>
<th></th>
<th>Not at all</th>
<th>Perfectly</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>0 1 2 3 4 5 6 7 8 9 10</td>
<td>Not applicable</td>
</tr>
<tr>
<td>P2</td>
<td>0 1 2 3 4 5 6 7 8 9 10</td>
<td>Not applicable</td>
</tr>
</tbody>
</table>

7. You are talking to someone in a place where there are a lot of echoes, such as a church or railway terminus building. Can you follow what the other person says?

<table>
<thead>
<tr>
<th></th>
<th>Not at all</th>
<th>Perfectly</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>0 1 2 3 4 5 6 7 8 9 10</td>
<td>Not applicable</td>
</tr>
<tr>
<td>P2</td>
<td>0 1 2 3 4 5 6 7 8 9 10</td>
<td>Not applicable</td>
</tr>
</tbody>
</table>

8. Can you have a conversation with someone when another person is speaking whose voice is the same pitch as the person you’re talking to?

<table>
<thead>
<tr>
<th></th>
<th>Not at all</th>
<th>Perfectly</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>0 1 2 3 4 5 6 7 8 9 10</td>
<td>Not applicable</td>
</tr>
<tr>
<td>P2</td>
<td>0 1 2 3 4 5 6 7 8 9 10</td>
<td>Not applicable</td>
</tr>
</tbody>
</table>
Speech Spatial Qualities (Part 1: Speech hearing)

<table>
<thead>
<tr>
<th>Question</th>
<th>P1</th>
<th>P2</th>
</tr>
</thead>
<tbody>
<tr>
<td>9. Can you have a conversation with someone when another person is</td>
<td>[Not applicable]</td>
<td>[Not applicable]</td>
</tr>
<tr>
<td>speaking whose voice is different in pitch from the person you're</td>
<td>123456789 1 0</td>
<td>123456789 1 0</td>
</tr>
<tr>
<td>talking to?</td>
<td>Not applicable</td>
<td>Not applicable</td>
</tr>
<tr>
<td></td>
<td>0 1 2 3 4 5 6 7 8 9 10</td>
<td>0 1 2 3 4 5 6 7 8 9 10</td>
</tr>
</tbody>
</table>

| 10. You are listening to someone talking to you, while at the same      | [Not applicable]                                                   | [Not applicable]                                                   |
| time trying to follow the news on TV. Can you follow what both people   | 123456789 1 0                                                     | 123456789 1 0                                                     |
| are saying?                                                             | Not applicable                                                     | Not applicable                                                     |
|                                                                        | 0 1 2 3 4 5 6 7 8 9 10                                               | 0 1 2 3 4 5 6 7 8 9 10                                               |

| 11. You are in conversation with one person in a room where there are   | [Not applicable]                                                   | [Not applicable]                                                   |
| many other people talking. Can you follow what the person you are      | 123456789 1 0                                                     | 123456789 1 0                                                     |
| talking to is saying?                                                  | Not applicable                                                     | Not applicable                                                     |
|                                                                        | 0 1 2 3 4 5 6 7 8 9 10                                               | 0 1 2 3 4 5 6 7 8 9 10                                               |

| 12. You are with a group and the conversation switches from one person  | [Not applicable]                                                   | [Not applicable]                                                   |
| to another. Can you easily follow the conversation without missing the | 123456789 1 0                                                     | 123456789 1 0                                                     |
| start of what each new speaker is saying?                              | Not applicable                                                     | Not applicable                                                     |
|                                                                        | 0 1 2 3 4 5 6 7 8 9 10                                               | 0 1 2 3 4 5 6 7 8 9 10                                               |
Speech Spatial Qualities (Part 1: Speech hearing)

<table>
<thead>
<tr>
<th>Question</th>
<th>P1</th>
<th>P2</th>
</tr>
</thead>
<tbody>
<tr>
<td>13. Can you easily have a conversation on the telephone?</td>
<td>Not applicable</td>
<td>Not applicable</td>
</tr>
<tr>
<td>14. You are listening to someone on the telephone and someone next to you starts talking. Can you follow what's being said by both speakers?</td>
<td>Not applicable</td>
<td>Not applicable</td>
</tr>
</tbody>
</table>

Overall Preference

<table>
<thead>
<tr>
<th>Preference</th>
<th>P1</th>
<th>P2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Which program do you prefer when listening in QUIET?</td>
<td>Not applicable</td>
<td>Not applicable</td>
</tr>
<tr>
<td>2. Which program do you prefer when listening in Noise?</td>
<td>Not applicable</td>
<td>Not applicable</td>
</tr>
</tbody>
</table>

Please comment about programs 1 and 2 – do you like the program or not, what you didn’t like about them, any differences in the way they respond to noises (for example traffic, party or constant noises such as fan or computer), sudden loud sounds, in quiet etc:

P1

P2

THANK YOU FOR COMPLETING THIS QUESTIONNAIRE
References


References


Related publications

The following papers have been published from the work in this thesis. Full text copies of each paper are also provided in this section.


Author/s:
HERSBACH, ADAM

Title:
Noise reduction for cochlear implants

Date:
2014

Persistent Link:
http://hdl.handle.net/11343/52813

File Description:
Noise reduction for cochlear implants