Investigating the effect of spatial separation on the detection of sounds in competition, by examining electrophysiological responses from the brainstem and auditory cortex

Nematollah Rouhbakhsh

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

June 2016

Department of Audiology and Speech Pathology
The University of Melbourne
Dedication

This thesis is dedicated to my patient family:

Fatemeh and Baran
Abstract

Human communication frequently takes place in noisy environments. In these environments, successful understanding of speech is dependent on an individual’s ability to extract and use spatial cues for separating speech from distracting noise. When speech and noise are separated spatially, the speech reception threshold (SRT) is reduced, and this is referred to in the literature as spatial release from masking (SRM). SRM is in most part due to acoustic cues arising from the differences in time and intensity of signals arriving at each of our two ears (i.e. interaural time (ITD) and interaural level (ILD) differences). ITDs and ILDs have in general been investigated through psychoacoustic studies. However, the electrophysiological correlates of these acoustic cues have only been investigated individually. For this reason, a novel experiment was designed to investigate the effect of spatial separation on the detection of target sound in competition with distractor stimuli in a more realistic experimental environment in which both ITD and ILD cues were present.

The primary aim of this thesis was to determine whether it is possible to identify a neural representation of SRM in the electrophysiological responses recorded from either the brainstem or auditory cortex, or both, using experimental stimuli conveying ILD and ITD cues. This research was conducted in two primary studies. The first study investigated whether the frequency-following response (FFR) in response to the fundamental frequency (F0) of a speech sound could be used to demonstrate SRM at different signal-to-noise ratios (SNRs), and what role of attentional mechanisms might play in spatial processing. FFRs were recorded in eighteen normally hearing participants. Participants were presented through headphones with a synthesized steady-state vowel /u/ with an F0 of 110 Hz and a 250 ms duration at 60 dB SPL. This vowel was labelled as the target stimulus. To be able to measure the effects of attention, a deviant stimulus was interspersed randomly throughout the target stimuli. To be able to measure the number of deviansthat occurred. It was assumed that, while identifying the number of deviant stimuli, the participant was actively listening to the stimuli. In the “non-attended” phase, participants were asked to ignore both the target and deviant stimuli, and any distractors.
The distractors were two continuous different stories spoken by the same speaker. Target, deviant and distractor stimuli were convolved with head-related transfer functions (HRTFs) to create two spatial conditions: the co-located condition with targets, deviants and distractors coming from 0° azimuth; and the separated condition with the targets and deviants at 0°, but with each distractor shifted to each side (± 90° azimuth). Three SNRs were considered (-5, -0, and 5 dB).

The amplitudes of the FFR in response to F0 were determined and analysed. The results of study 1 revealed a significant effect of spatial separation. The effect of spatial separation was found only at the lower SNR. Spatially separating maskers from the target stimuli resulted in a significant larger amplitude of the FFR in response to the target F0. The spatial advantage obtained objectively was equivalent to an SNR increase of 3.3 dB. A significant effect for attention was found when participants actively focused on the target, as demonstrated by larger FFR amplitudes. However, no significant interactions were found between spatial separation and the level of attention.

The findings of the first study suggest that binaural processing relevant to SRM may be reflected by phase locked neural activity in the brainstem. However, this objective measure may only be noticeable in relatively noisy environments. Furthermore, SRM may start early in the central auditory pathways regardless of one attending to the target stimuli or not. This last observation means that - although this thesis focuses on adults - an extrapolation potentially could be made towards the use with younger individuals, however with consideration of their brain differences with adults and the AEPs evoked from those brains. The lack of dependence on attention might be beneficial in investigating SRM in this population, where it is difficult to keep attention and one has to rely on objective techniques that do not require attending to the target stimulus. Conversely, the lack of interaction with attention may mean that the mechanism responsible for the objective results may be different from the mechanism primarily responsible for SRM.

To identify whether objective markers of SRM can be recorded in either the brainstem or cortex (or both), a second study was conducted. In the second study, auditory brainstem responses (ABRs), FFRs, and cortical auditory evoked potentials (CAEPs) were recorded simultaneously from thirteen normally hearing adults in response to 200 target stimulus blocks. Each target stimulus block comprised of a series of 11 tone complexes (TCs), with each TC having a specific F0 and a duration of 30 ms, separated by a 30 ms
interstimulus interval (ISI), resulting in a target stimulus block with a total duration of 630 ms. The blocks were repeated every 1200 ms. Two different target stimulus block paradigms were considered; flat and staircase. The ‘flat’ blocks had TCs with a constant fundamental frequency \( F_0 \) of 325 Hz (and harmonics up to 6 kHz). In the ‘staircase’ blocks, the fundamental frequency of each TCs was reduced in steps of 30 Hz from 475 to 175 Hz, again with harmonics up to 6 kHz.

ABRs were recorded to the onsets of the 30-ms TCs. FFRs were recorded in response to the F0s of the TCs, and CAEPs to the onsets of the target stimulus blocks. The distractor blocks, in contrast, comprised of blocks of TCs that were similar in number and duration to the target stimuli, but randomized in their F0 distribution from 100 to 550 Hz and jittered in time (+/- 15 ms) around the onset of the target TC. Both target and distractor stimuli were convolved with head-related transfer functions (HRTFs) and presented under headphones. The target stimuli were presented at 0° azimuth. The distractors were co-located (at 0° azimuth) and spatially separated (at ±90° azimuth) from the targets. The targets were presented at SNRs of -5, 0, 5, 10 and 15 dB SNR, and at 60 dB SPL.

After extraction and analysis of ABR amplitudes and latencies, and FFR amplitudes, the results of the second study revealed a significant effect of SRM as seen in a decrease in ABR latency for both flat and staircase target stimuli when spatially separating maskers from the target. FFR amplitude (only measured with the flat stimuli) was significantly larger in the separated condition, and a significant decrease in CAEP latencies (for the staircase stimuli) was found, but only at the lowest tested SNR of -5 dB. These results, particularly the FFR, confirmed the results obtained in the first study, i.e. separating distractors from the target, regardless of the type of stimulus being used, resulted in enhancing FFR F0 amplitude. However, due to noisy data, the observations at the cortical level need to be confirmed in a follow-up study. The spatial advantage was equivalent to a SNR increase of 4.3 dB for FFR amplitude (for the flat stimuli), and 13.8 dB for ABR latency, 11.2 dB for CAEP P1 and 19.9 dB for CAEP N1 latencies (for the staircase stimuli).

The findings of the second study suggest that it is possible objectively to record SRM in both the brainstem and auditory cortex simultaneously at lower SNRs. This suggests that the central auditory system is able to squelch background noise via processing of spatial information, and that this capacity is higher in more challenging listening environments.
Taken together, the results from the first and second studies suggest that it is feasible to use electrophysiological measures as a means of investigating the central auditory mechanisms, which contribute to SRM in the brainstem and cortex simultaneously. It is speculated that SRM occurs mainly at the level of the brainstem and is present at -5 dB SNR (i.e. difficult listening environments). The finding that SRM was primarily at lower SNRs is in reality not a clinical concern, as lower SNRs represent the environments in which SRM is generally found to be beneficial for the listener. Potential applications may be found in developing an objective detection test for spatial processing disorder (SPD), a condition in which normal-hearing individuals are unable to exploit the binaural mechanism of SRM when listening in noisy environments, i.e., a deficiency in selectively attending to target sounds, which are not spatially co-located with distractor sounds. Further studies are needed to investigate the effects of attention and SRM on brainstem and cortical responses in different populations including children, elderly, and people with SPD.
Declaration

This is to certify that:

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

-------------------------------
Nematollah Rouhbakhsh
Preface

This thesis is composed of my original work and contains no material previously published or written by another person except where due reference has been made. I have clearly stated the contribution of others to my thesis as a whole, including statistical assistance, data analysis, and other original research work used or reported in my thesis.

The content of my thesis is the result of work I have carried out since the commencement of my research higher degree candidature, and does not include work that has been submitted to qualify for the award of any other degree or diploma in any university or tertiary institution.

I acknowledge that an electronic copy of my thesis must be lodged with the University Library and, subject to Award Rules of the University of Melbourne, immediately made available for research and study in accordance with the Copyright Act.

I acknowledge that this research was financially supported by the Hearing Cooperative Research Centre, established under the Australian Government’s Cooperative Research Centres Program. I acknowledge that the copyright of all material contained in my thesis resides with the Hearing CRC as copyright holder[s] of that material. In addition, I acknowledge the generous support of the Tehran University of Medical Sciences (TUMS).

The study in Chapter 3 was developed in collaboration with my supervisors, and with Dr Jörg Buchholz at the National Acoustic Laboratories (NAL) and Professor Ananthanarayan Krishnan (Purdue University). Chapter 4 was developed in collaboration with my supervisors and Dr Jörg Buchholz. Mark Seeto provided statistical support. In addition to my supervisors, I am indebted to Dr Wendy Noble (Chapters 1, 2, and 5), Teagan Young (Chapter 2) and Mark Peters (Chapter 5 and a part of Chapter 2), who proofread parts of this thesis.
Acknowledgements

First of all, I am grateful to the almighty Allah (God) for enabling me to complete this thesis.

Without the financial support of the Hearing Cooperative Research Centre and in particular the Tehran University of Medical Sciences (TUMS), I could not have undertaken this thesis.

My most cordial thanks are due to the National Acoustic Laboratories (NAL) for the use of its facilities, which made this project feasible.

I would like to express the greatest appreciations to my supervisors:

- Doctor Bram Van Dun, who was always ready to help me with his encouraging, insightful and sometimes critical feedback. I am extremely grateful and indebted to him for his expert advice, sincere and valuable guidance, and encouragement. Thanks Bram.

- Professor Harvey Dillon, Director of the National Acoustic Laboratories (NAL), who introduced me to a new field of hearing science, i.e. spatial processing, and for providing me with all the necessary facilities and with his opportune pointers as well. Your insight and guidance along the way have been invaluable.

- Professor Robert Cowan, CEO of the Hearing CRC, who assisted both scientifically and executively and for allowing me to fulfil my dream of being a student at the University of Melbourne. I am extremely grateful for your assistance and suggestions throughout my project.

I would like to offer my sincere thanks to my Advisory Committee members, Associate Professor Gary Rance and Dr Richard Van Hoesel, who helped me through their guidance and assistance.

I also take this opportunity to express my thanks to Professor Richard Dowell, the Head of the Department of Audiology and Speech Pathology, The University of Melbourne, for his kind help.

My sincere gratitude is extended to the Australian Hearing Human Research Ethics Committee for sharing their time, commitment, and positive insights into this project.
I wish to express my thanks to Dr Jörg Buchholz for his technical assistance and ideas on stimulus development and physiological mechanisms.

I would like to express my special gratitude to Professor Ananthanarayan Krishnan for his ideas, for the development of the protocol, and for help in reviewing Chapter 3 of this thesis.

I thank Mark Seeto for his ready, real-time answers to my puzzling statistical questions.

I have to express my special thanks to my proof-readers Dr Wendy Noble (for editing Chapters 1, 2, and 5) Teagan Young (for editing Chapter 2) and Mark Peters (for editing Chapter 5 and part of Chapter 2) for doing a great job on my thesis.

I take this opportunity to record my sincere thanks to the entire NAL staff, who directly or indirectly, has lent a helping hand in this project with their generosity of time and effort. In particular, I want to especially thank Vivian Fabricatorian, Greg Stewart, and John Seymour for their exemplary assistance.

I have to express my special thanks to Dr Jane Sewell and Amanda Campbell from the Hearing CRC for their incredible assistance and efficiency in getting things done.

I am also indebted to all of the subjects who volunteered to participate in this project.

Finally, to you, my exceptional wife, Fatemeh, I express my deepest thanks and love for all your support and sincerity. I apologize that our time in Sydney has been dominated by my research pursuits. Thank you for your patience. Undoubtedly, without you, completing this project would not have been possible.

Grieving the loss of my father during the second year of my candidacy was overwhelming for me. May Allah rest his soul in peace, bless him with his mercy, and grant him the sublime place in heaven. Amen.
Contents

Dedication ........................................................................................................................................ iii
Abstract ............................................................................................................................................... v
Declaration ......................................................................................................................................... ix
Preface ................................................................................................................................................ xi
Acknowledgements ............................................................................................................................ xiii
Contents .............................................................................................................................................. xv
Table of figures ................................................................................................................................... xix
Table of tables .................................................................................................................................... xxiii
List of Abbreviations ............................................................................................................................ xxv

Chapter 1. Introduction ............................................................................................................................... 1
1.1. Background and significance ............................................................................................................. 2
1.2. Objectives and research questions .................................................................................................. 6
1.3. Thesis outline .................................................................................................................................... 7

Chapter 2. Spatial processing and its disorder: Neurophysiological and electrophysiological approaches .......................................................................................................................... 10
2.1. Preface ............................................................................................................................................. 10
2.2. Introduction ..................................................................................................................................... 10
2.3. Central auditory processing (CAP) .................................................................................................. 12
2.3.1. Central auditory processing disorder ((C)APD) .......................................................................... 13
2.3.1.1. Spatial processing ....................................................................................................................... 16
2.4. The auditory system and its relevance for spatial processing .......................................................... 20
2.4.1. The peripheral auditory system .................................................................................................... 20
2.4.1.1. Outer ear ..................................................................................................................................... 21
2.4.1.2. Middle ear ................................................................................................................................. 21
2.4.1.3. Inner ear ..................................................................................................................................... 23
2.4.1.4. Auditory nerve (AN) ................................................................................................................. 24
2.4.2. The central auditory system ........................................................................................................ 26
2.4.2.1. Cochlear nucleus (CN) .............................................................................................................. 26
2.4.2.2. Superior olivary complex (SOC) ............................................................................................... 28
2.4.2.2.1. Medial superior olive (MSO) ................................................................................................. 28
2.4.2.2.2. Lateral superior olive (LSO) ................................................................................................. 29
2.4.2.3. Lateral lemniscus (LL) .............................................................................................................. 32
2.4.2.4. Inferior colliculus (IC) .............................................................................................................. 34
2.4.2.5. Medial geniculate body (MGB) ................................................................................................. 38
2.4.2.6. Auditory cortex (AC) ............................................................................................................... 40
2.5. Auditory-evoked potentials (AEPs) and their potential for assessing spatial processing .................. 45
2.5.1. Electrocochleography (ECochG) ................................................................................................. 47
2.5.2. The auditory brainstem response (ABR) ..................................................................................... 48
2.5.2.1. Possible generators underpinning the ABR .............................................................................. 49
Chapter 3. Spatial release from masking (SRM): Human frequency-following response (FFR) correlation

3.1. Preface ................................................................. 72
3.2. Introduction .......................................................... 72
3.3. Materials and methods .................................................. 77
  3.3.1. Participants ......................................................... 77
  3.3.2. Procedure ......................................................... 77
  3.3.3. Stimuli ............................................................ 78
  3.3.4. EEG recording .................................................... 82
  3.3.5. Data analysis ....................................................... 83
  3.3.6. Statistical analysis ............................................... 84
3.4. Results ................................................................. 86
  3.4.1. Evidence for SRM in FFR F₀ amplitude .......................... 86
  3.4.2. Effects of Stimulus SNR on FFR F₀ amplitude ................. 88
    3.4.2.1. Main effects of stimulus SNR ................................ 88
    3.4.2.2. Interaction effects of SRM and SNR ..................... 88
  3.4.3. Effects of attention on FFR F₀ amplitude .................... 89
    3.4.3.1. Main effects of attention ................................ 89
    3.4.3.2. Interaction effects of attention and SRM .............. 90
    3.4.3.3. Interaction effects of attention and SNR ............... 91
  3.4.4. Effect of channel on evaluated parameters ................... 92
  3.4.5. Magnitude of effects ........................................... 92
3.5. Discussion ............................................................ 93
  3.5.1. Binaural spatial release from masking at the brainstem level 94
  3.5.2. Frequency-following response correlates with spatial release from masking at lower SNRs ........................... 97
  3.5.3. Attention modulates FFR F₀ amplitude but might not modulate SRM at the level of the brainstem .................. 98
  3.5.4. Effects of channels .......................................... 100
Chapter 4. Investigating spatial processing through the simultaneous recording of the auditory brainstem response (ABR), frequency-following response (FFR) and cortical auditory evoked response (CAEP) ........................................95
4.1. Preface .........................................................................................................................104
4.2. Introduction ..................................................................................................................104
4.3. Materials and methods .................................................................................................107
4.3.1. Participants ............................................................................................................107
4.3.2. Stimuli ......................................................................................................................107
4.3.3. Procedure ...............................................................................................................113
4.3.3.1. Behavioural evaluation .......................................................................................113
4.3.3.2. Electrophysiological evaluation .........................................................................114
4.3.4. Data analysis ..........................................................................................................115
4.3.5. Statistical analysis ...................................................................................................118
4.4. Results .........................................................................................................................119
4.4.1. Behavioural results (LiSN-T) ................................................................................120
4.4.2. Auditory brainstem response .................................................................................120
4.4.2.1. Wave V amplitude ..............................................................................................120
4.4.2.1.1. Main effect of SRM on wave V amplitude .......................................................122
4.4.2.1.2. Main effect of SNR on wave V amplitude .......................................................123
4.4.2.1.3. Main effect of target frequency paradigm on wave V amplitude ...............123
4.4.2.1.4. Interaction effect between SRM and SNR on wave V amplitude ...............123
4.4.2.1.5. Interaction effect between SRM and target frequency paradigm on wave V amplitude ........................................................................................................123
4.4.2.2. Wave V latency ....................................................................................................124
4.4.2.2.1. Main effect of SRM on wave V latency .........................................................125
4.4.2.2.2. Main effect of SNR on wave V latency .........................................................126
4.4.2.2.3. Main effect of target frequency paradigm on wave V latency .................126
4.4.2.2.4. Interaction effect between SRM and SNR on wave V latency .................126
4.4.2.2.5. Interaction effect between SRM and target frequency paradigm on wave V latency ........................................................................................................127
4.4.3. Frequency-following response (FFR) ....................................................................127
4.4.3.1. Main effect of SRM on F0 amplitude ...............................................................127
4.4.3.2. Main effect of SNR on F0 amplitude ...............................................................128
4.4.3.3. Interaction effect between SRM and SNR on F0 amplitude ............................129
4.4.4. Cortical auditory evoked potential (CAEP) ..............................................................129
4.4.4.1. P1 amplitude ........................................................................................................130
4.4.4.1.1. Main effect of SRM on P1 amplitude ............................................................130
4.4.4.1.2. Main effect of SNR on P1 amplitude ............................................................130
4.4.4.1.3. Main effect of target frequency paradigm on P1 amplitude .......................130
4.4.4.1.4. Interaction effects between SRM and SNR on P1 amplitude .......................130
4.4.4.1.5. Interaction effects between SRM and target frequency paradigm on P1 amplitude ........................................................................................................131
4.4.4.2. P1 latency ............................................................................................................131
4.4.4.2.1 Main effect of SRM on P1 latency ...............................................................131
4.4.4.2.2. Main effect of stimulus SNR on P1 latency ................................................132
4.4.4.2.3. Main effect of target frequency paradigm on P1 latency ............................132
4.4.4.2.4. Interaction effects between SRM and SNR on P1 latency ............................132
Chapter 5. Summary, limitations, clinical significance, and future research .....................................141

5.1. Preface .........................................................................................................................156
5.2. Summary .....................................................................................................................156
5.2.1. Introduction .............................................................................................................156
5.2.2. Neurophysiological and electrophysiological approaches to spatial processing: theoretical bases .................................................................157
5.2.3. Spatial release from masking: Human frequency-following response correlation ........................................................................................................159
5.2.4. Investigating spatial processing through the simultaneous recording of the auditory brainstem response, frequency following response, and cortical auditory evoked potential .....................................................................................160
5.3. Objective and research questions .............................................................................162
5.4. Limitations ..................................................................................................................162
5.5. Clinical significance .................................................................................................165
5.6. Directions for future research ..................................................................................166
5.7. Conclusion ..................................................................................................................168

References ..........................................................................................................................169
Table of figures

Figure 1. The structure of the peripheral human ear................................................................. 21
Figure 2. The head-related transfer function (HRTF)................................................................. 22
Figure 3. Cross-section of the cochlea.......................................................................................... 23
Figure 4. Schematic of the olivocochlear bundle (OCB) to the right cochlear of a cat............. 26
Figure 5. The original Jeffress (1948) delay coincidence model for the right side of the brainstem... 30
Figure 6. Responses of LSO cell to 200 ms broadband noise introduced from five different azimuths in binaural and monaural ipsilateral ear only................................................................. 31
Figure 7. Different connections of the LL..................................................................................... 33
Figure 8. One example of a neural cell in the ventral part of the medial geniculate body (MGB)...... 39
Figure 9. Cortical and subcortical connections of the primate auditory system......................... 41
Figure 10. The three schematic graphs illustrate idealized spike rate versus ILD functions for different neurons......................................................................................................................... 43
Figure 11. Spike patterns of a single unit in the core auditory cortex............................................. 44
Figure 12. Schematic diagram of instrumentation applied when recording AEPs and the related factors that affect this assessment................................................................................... 46
Figure 13. Classification of auditory evoked potentials based on their latency............................. 47
Figure 14. Example of an electrocochleogram waveform............................................................. 48
Figure 15. Example of an ABR waveforms and suggested generators for major peaks of the auditory brainstem response..................................................................................................... 49
Figure 16. An example of a FFR waveform in time and frequency domains............................... 53
Figure 17. Example of a MLR waveform ....................................................................................... 59
Figure 18. Example of a CAEP waveform...................................................................................... 60
Figure 19. The reversal trend in polarity of CAEP components over the regions of the generator sites of the scalp.................................................................................................................... 61
Figure 20. Example of a P300 waveform....................................................................................... 65
Figure 21. The spectrogram, and spectra of /u/ target stimulus combined with two-talker discourses for two channels in co-located condition, and separated condition........................................... 80
Figure 22. Distribution of local SNRs of -5, 0, and 5 dB. Distribution of local SNRs for low, medium, and high ranges.................................................................................................................. 82
Figure 23. Means and standard errors of response SNR for each component of the FFR, averaged across recording channels........................................................................................................ 85
Figure 24. Means and standard errors of response SNR of FFR in recording channels, averaged across response components........................................................................................................ 85
Figure 25. Means and standard errors of FFR F0 amplitude in spatially co-located and separated conditions as a function of stimulus SNR, averaged across the two attention conditions and channels one and three.............................................................. 87
Figure 26. Grand averages of 18 participants showing the FFR in the frequency domain for low, medium and high SNRs, collapsed across 2 attention and 2 spatial conditions, and channels 1 and 3................................................................................................................................... 88
Figure 27. Grand averages of FFR F0 amplitude over 18 participants of the response spectra for stimulus local SNRs: low, medium and high in co-located and separated spatial condition........ 89
Figure 28. Means and standard errors of FFR F0 amplitude as a function of attention, averaged across channels one and three, both spatial conditions, and all SNRs..................................................... 90
Figure 29. Means and standard errors of FFR F0 amplitude in spatially co-located and separated conditions, averaged across channels one and three and local SNRs............................................................... 90
Figure 30. Grand averages of FFR F0 amplitude over 18 participants of the response spectra for stimulus local SNRs: low, medium and high in attention and non-attention conditions, averaged across two spatial conditions and channels........................................................................................................ 91
Figure 31. Means and standard errors of F0 amplitude in attention conditions as a function of stimulus SNR, averaged across spatial conditions and channels one and three. ........................................... 91
Figure 32. Means and standard errors of F0 amplitude in channel one and three averaged across spatial conditions, SNR and attention conditions. ................................................................. 92
Figure 33. Eleven 30-ms long target tone complexes are presented together with jittered 30-ms long distractors. These target-distractor combinations are then followed by nine distractors. The flat targets have a constant F0 of 325 Hz. The downward staircase targets have F0s that drop from 475 to 175 Hz in steps of 30 Hz. The F0s of the distractors are randomised between 100 and 550 Hz. ................................................................................................................................. 109
Figure 34. The waveform, the spectrogram and spectra of TC target stimulus in flat frequency paradigm. The waveforms, spectrograms and spectra of TC discourses in co-located left and right channel, separated left and right channel................................................................. 111
Figure 35. The waveform, the spectrogram and spectra of TC target stimulus in staircase frequency paradigm. The waveforms, spectrograms and spectra of TC discourses in co-located left and right channel, separated left and right channel................................................................. 112
Figure 36. Grand average and individual traces over all 13 subjects of the 15 dB SNR co-located conditions................................................................................................................................. 117
Figure 37. Grand averages from 13 subjects for -5, 0, 5, 10 and 15 dB SNR for flat stimulus co-located , flat stimulus separated, staircase stimulus co-located and staircase stimulus separated. ..... 119
Figure 38. Grand averages waveforms of FFR F0 amplitude overlayed13participantsin time domain at 5 tested SNRs in co-located and separated spatial conditions for flat frequency paradigm. ABR wave V is considered as the first positive of the first FFR positive waveform................. 121
Figure 39. Means and standard errors of ABR wave V amplitude in spatially co-located and separated conditions as a function of target SNR, averaged across test and re-test when using a flat target frequency paradigm. ................................................................................................................................. 121
Figure 40. Grand averages waveforms of FFR F0 amplitude overlayed13participantsin time domain at 5 tested SNRs in co-located and separated spatial conditions for staircase frequency paradigm. ABR wave V is considered as the first positive of the first FFR positive waveform. ................. 122
Figure 41. Means and standard errors of ABR wave V amplitude in spatially co-located and separated conditions as a function of target SNR, averaged across test and re-test when using a staircase target frequency paradigm ................................................................. 122
Figure 42. Means and standard errors of ABR latency in spatially co-located and separated conditions as a function of target SNR, averaged across test and re-test when using a flat target frequency paradigm. ................................................................................................................................. 125
Figure 43. Means and standard errors of ABR latency in spatially co-located and separated conditions as a function of target SNR, averaged across test and re-test when using a staircase target frequency paradigm................................................................. 125
Figure 44. Grand averages waveforms of FFR F0 amplitude overlayed 13 participantsin frequency domain at 5 tested SNRs in co-located and separated spatial conditions for flat frequency paradigm. ................................................................................................................................. 128
Figure 45. Means and standard errors of FFR F0 amplitude in spatially co-located and separated conditions as a function of stimulus SNR when using a flat target frequency paradigm. 128
Figure 46. Means and standard errors of P1 amplitude in spatially co-located and separated conditions................................................................................................................................. 131
Figure 47. Means and standard errors of P1 latency in spatially co-located and separated conditions132
Figure 48. Means and standard errors of N1 amplitude in spatially co-located and separated conditions as a function of stimulus SNRs, averaged across the test and re-test in a flat and staircase frequency paradigms. ................................................................................................................................. 135
Figure 49. Means and standard errors of N1 latency in spatially co-located and separated conditions as a function of stimulus SNRs, averaged across the test and re-test in a flat and staircase frequency paradigm. ................................................................................................................................. 136

xx
Figure 50. Scatterplot of the correlation between LiSN-T and spatial advantage of FFR F0 amplitude at -5 dB SNR, ABR latency in flat and staircase frequency paradigms across all SNRs. ............ 139
Figure 51. Scatterplot of the correlation between the LiSN-T and spatial advantage of P1 and N1 latency at -5 dB SNR in flat and staircase frequency paradigms ........................................... 140
Table of tables

Table 1. Summary of the statistical results of a 4-way 2x3x2x2 repeated measures ANOVA with SRM, SNR, channel and attention condition as repeated-measures variables........................................87
Table 2. Results of repeated measures analysis of amplitude and latency of the wave V ABR, FFR F0 amplitude, and amplitudes and latencies of the CAEP components P1 and N1 for SNR, SRM and interaction between SRM and SNR using flat and staircase frequency paradigms (Stim). .......124
Table 3. Absolute benefit of SRM (in dB) for all significant measures in this study. CAEP results are provisional. The restrictive conditions indicate where the significant result was obtained. .....141
Table 4. Statistically significant spatial advantages (in dB) of FFR F0 amplitude, ABR wave V latency, CAEP N1 and P1 latency, and behavioural tests LISN-S and LISN-T .................................................164
**List of Abbreviations**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAF</td>
<td>Anterior auditory field</td>
</tr>
<tr>
<td>ABR</td>
<td>Auditory brainstem response</td>
</tr>
<tr>
<td>AC</td>
<td>Auditory cortex</td>
</tr>
<tr>
<td>AHREC</td>
<td>Australian Hearing Human Research Ethics Committee</td>
</tr>
<tr>
<td>AI</td>
<td>Primary auditory cortex</td>
</tr>
<tr>
<td>AII</td>
<td>Secondary auditory cortex</td>
</tr>
<tr>
<td>AIII</td>
<td>Tertiary auditory cortex</td>
</tr>
<tr>
<td>AL</td>
<td>Anterior lateral</td>
</tr>
<tr>
<td>AM</td>
<td>Amplitude modulation</td>
</tr>
<tr>
<td>ANOVA</td>
<td>Analysis of variance</td>
</tr>
<tr>
<td>AL</td>
<td>Anterolateral area</td>
</tr>
<tr>
<td>AP</td>
<td>Action potential</td>
</tr>
<tr>
<td>AVCN</td>
<td>Anteroventral cochlear nucleus</td>
</tr>
<tr>
<td>BIC</td>
<td>Binaural interaction component</td>
</tr>
<tr>
<td>BMLD</td>
<td>Binaural masking level difference</td>
</tr>
<tr>
<td>CAEP</td>
<td>Cortical auditory evoked potential</td>
</tr>
<tr>
<td>CANS</td>
<td>Central auditory nervous system</td>
</tr>
<tr>
<td>CAP</td>
<td>Central auditory processing</td>
</tr>
<tr>
<td>(C)APD</td>
<td>Central auditory processing disorder</td>
</tr>
<tr>
<td>CF</td>
<td>Characteristic frequency</td>
</tr>
<tr>
<td>CL</td>
<td>Caudolateral area</td>
</tr>
<tr>
<td>CCOB</td>
<td>Crossed olivocochlear bundle</td>
</tr>
<tr>
<td>CN</td>
<td>Cochlear nuclei</td>
</tr>
<tr>
<td>CM</td>
<td>Caudomedial area</td>
</tr>
<tr>
<td>COM</td>
<td>Chronic otitis media</td>
</tr>
<tr>
<td>CPB</td>
<td>Caudal parabelt</td>
</tr>
<tr>
<td>dB</td>
<td>Decibel</td>
</tr>
<tr>
<td>dB HL</td>
<td>Decibels hearing level</td>
</tr>
<tr>
<td>dB SPL</td>
<td>Decibels sound pressure level</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>DCN</td>
<td>Dorsal cochlear nucleus</td>
</tr>
<tr>
<td>DDT</td>
<td>Dichotic digit test</td>
</tr>
<tr>
<td>DNLL</td>
<td>Dorsal nuclei of lateral lemniscus</td>
</tr>
<tr>
<td>DRR</td>
<td>Direct-to-reverberant energy ratio</td>
</tr>
<tr>
<td>EE</td>
<td>Excitatory-excitatory</td>
</tr>
<tr>
<td>EEG</td>
<td>Electroencephalogram</td>
</tr>
<tr>
<td>EI</td>
<td>Excitatory-Inhibitory</td>
</tr>
<tr>
<td>ERP</td>
<td>Evoked response potential</td>
</tr>
<tr>
<td>FFR</td>
<td>Frequency following response</td>
</tr>
<tr>
<td>FRA</td>
<td>Frequency response area</td>
</tr>
<tr>
<td>GABA</td>
<td>Gamma-aminobutyric acid</td>
</tr>
<tr>
<td>GIN</td>
<td>Gaps-in-noise test</td>
</tr>
<tr>
<td>HRTF</td>
<td>Head related transfer function</td>
</tr>
<tr>
<td>HSR</td>
<td>High spontaneous rate</td>
</tr>
<tr>
<td>IC</td>
<td>Inferior colliculus</td>
</tr>
<tr>
<td>ICC</td>
<td>Inferior colliculus, central nucleus</td>
</tr>
<tr>
<td>ICD</td>
<td>Inferior colliculus, dorsal cortex</td>
</tr>
<tr>
<td>ICX</td>
<td>Inferior colliculus, external cortex</td>
</tr>
<tr>
<td>IE</td>
<td>Inhibitory-excitatory</td>
</tr>
<tr>
<td>IHC</td>
<td>Inner hair cells</td>
</tr>
<tr>
<td>IHS</td>
<td>Intelligent Hearing Systems</td>
</tr>
<tr>
<td>ILD</td>
<td>Interaural intensity difference</td>
</tr>
<tr>
<td>IAS</td>
<td>Intermediate acoustic stria</td>
</tr>
<tr>
<td>ISD</td>
<td>Interaural spectral difference</td>
</tr>
<tr>
<td>ITD</td>
<td>Interaural time difference</td>
</tr>
<tr>
<td>kHz</td>
<td>kilohertz</td>
</tr>
<tr>
<td>Lim</td>
<td>Limitans nucleus</td>
</tr>
<tr>
<td>LiSN-S</td>
<td>Listening in Spatialised Noise – Sentences test</td>
</tr>
<tr>
<td>LiSN-T</td>
<td>Listening in Spatialised Noise – Tone test</td>
</tr>
<tr>
<td>LL</td>
<td>Lateral lemniscus</td>
</tr>
<tr>
<td>LOC</td>
<td>Lateral olivary complex</td>
</tr>
<tr>
<td>LP</td>
<td>Learning problem</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>LSO</td>
<td>Lateral superior olive</td>
</tr>
<tr>
<td>LSR</td>
<td>Low spontaneous rate</td>
</tr>
<tr>
<td>MEG</td>
<td>Magnetoencephalography</td>
</tr>
<tr>
<td>MGB</td>
<td>Medial geniculate body</td>
</tr>
<tr>
<td>MGd</td>
<td>Dorsal nucleus of the medial geniculate complex</td>
</tr>
<tr>
<td>MGm</td>
<td>Medial/magnocellular nucleus of the medial geniculate complex</td>
</tr>
<tr>
<td>MGv</td>
<td>Ventral nucleus of the medial geniculate complex</td>
</tr>
<tr>
<td>ML</td>
<td>Middle lateral area</td>
</tr>
<tr>
<td>MLD</td>
<td>Masking level difference</td>
</tr>
<tr>
<td>MM</td>
<td>Medial-medial</td>
</tr>
<tr>
<td>MMN</td>
<td>Mismatch negativity</td>
</tr>
<tr>
<td>MOC</td>
<td>Medial olivary complex</td>
</tr>
<tr>
<td>MSO</td>
<td>Medial superior olive</td>
</tr>
<tr>
<td>NH</td>
<td>Normal hearing</td>
</tr>
<tr>
<td>OAE</td>
<td>Otoacoustic emission</td>
</tr>
<tr>
<td>OCB</td>
<td>Olivocochlear bundle</td>
</tr>
<tr>
<td>OHC</td>
<td>Outer hair cell</td>
</tr>
<tr>
<td>RA</td>
<td>Reduced audibility</td>
</tr>
<tr>
<td>PAF</td>
<td>Posterior auditory field</td>
</tr>
<tr>
<td>PCN</td>
<td>Posteroventral cochlear nucleus</td>
</tr>
<tr>
<td>PM</td>
<td>Medial pulvinar nucleus</td>
</tr>
<tr>
<td>R</td>
<td>Rostral area (Rostral primary-like area)</td>
</tr>
<tr>
<td>RGDT</td>
<td>Random gap detection test</td>
</tr>
<tr>
<td>RM</td>
<td>Rostromedial area</td>
</tr>
<tr>
<td>RPB</td>
<td>Rostral parabelt</td>
</tr>
<tr>
<td>RT</td>
<td>Rostrotemporal primary-like cortex</td>
</tr>
<tr>
<td>RTM</td>
<td>Medial rostrotemporal area</td>
</tr>
<tr>
<td>RTL</td>
<td>Lateral rostrotemporal area</td>
</tr>
<tr>
<td>Sg</td>
<td>Suprageniculate nucleus</td>
</tr>
<tr>
<td>SIN</td>
<td>Speech in noise</td>
</tr>
<tr>
<td>SNR</td>
<td>Signal-to-noise ratio</td>
</tr>
<tr>
<td>SPD</td>
<td>Spatial processing disorder</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td>SRF</td>
<td>Spatial receptive field</td>
</tr>
<tr>
<td>SRM</td>
<td>Spatial release from masking</td>
</tr>
<tr>
<td>SRT</td>
<td>Speech reception threshold</td>
</tr>
<tr>
<td>SSW</td>
<td>Staggered spondaic word test</td>
</tr>
<tr>
<td>STG</td>
<td>Superior temporal gyrus</td>
</tr>
<tr>
<td>STR</td>
<td>Stochastic resonance</td>
</tr>
<tr>
<td>TB</td>
<td>Trapezoid body</td>
</tr>
<tr>
<td>TC</td>
<td>Tone-complex</td>
</tr>
<tr>
<td>Tpt</td>
<td>Temporoparietal area</td>
</tr>
<tr>
<td>TS1, 2</td>
<td>Superior temporal areas 1 and 2</td>
</tr>
<tr>
<td>UCOB</td>
<td>Uncrossed olivocochlear bundle</td>
</tr>
<tr>
<td>VNLL</td>
<td>Ventral nuclei of lateral lemniscus</td>
</tr>
</tbody>
</table>
Chapter 1. Introduction
1.1. **Background and significance**

In everyday listening environments, our ears receive acoustic signals from a variety of sound sources that may occur in part or entirely at the same time. These signals interleave and combine into one complex acoustic environment, which reaches our ears. The activity of following a conversation in such a complex environment is made increasingly difficult as the number of acoustic sources in the acoustic environment increases.

In normally hearing individuals, one of the functions of activity in the central auditory nervous system (CANS) is to separate signals embedded within this complex acoustic environment into the respective source components. Source separation is based in part on perception of differences in the respective spatial cues, primarily inter-aural differences in the intensity and timing of individual arriving signals Bregman (1990), Carlyon, 2004, Shamma and Micheyl, 2010).

The central auditory nervous system primarily uses two strategies for locating signals in the horizontal plane within a complex acoustic environment. First, it utilizes the differences in time of arrival of signals at the two external ears on opposite sides of the head (measured as inter-aural time differences (ITDs)). Second, it calculates the changes in the intensity of signals between the two ears (measured as inter-aural intensity or level difference (ILD) cues). Whereas ITDs occur primarily from the time-of-arrival differences resulting from the greater distance travelled by a signal to reach the distal ear, ILDs occur primarily from diffraction around the head, and in particular the attenuation of sound levels at the distal side of the head furthest from the acoustic source. The level of attenuation will differ dependent on the specific frequency components of the signal, with greater attenuation in the higher frequencies (from about 2 kHz and up). The ILDs cause SNR differences between the signal at the two ears when the target signal and competing sounds come from different source directions (Marrone et al., 2008).

The ability of the CANS to analyse acoustic environments degrades as the complexity of the environment increases (Schneider et al., 2007). The complexity of acoustic environments is either due to overlap of competing signals in time and frequency with target signals, which is known as energetic masking (EM), or is due to an increase in similarity of content between target signals and competing signals, which is known as informational masking (IM) (Kidd et al., 1994).
Central auditory processing (CAP) refers to the neurological activities underlying the perceptual processing of auditory information. These neurological processes give rise to electrophysiological potentials arising from different locations within the CANS that can be recorded using a range of electrophysiological test measures (ASHA, 2005). The recording of electrophysiological potentials, arising in response to auditory processing of acoustic information, termed auditory evoked potentials (AEPs), presents a window that enables researchers to evaluate a variety of related auditory behaviours such as sound localization and lateralization, as well as auditory perception in situations of competing or degraded acoustic information.

Deficits in the perceptual processing of acoustic information within the CANS are generally referred to as central auditory processing disorders ((C)APD). Any deficit in the perceptual processing of auditory information must be reflected in changes to the underpinning neurobiological activity, which potentially can be then detected through changes in the characteristics of the electrophysiological auditory potentials that can be recorded from different locations within the CANS (ASHA, 2005). Auditory evoked potentials can provide valuable information in the evaluation of (C)APD as they can be evoked primarily from different locations within the CANS, for example the auditory brainstem response (ABR) as contrasted with higher brain level cortical auditory evoked potentials (CAEPs) (Jerger and Musiek, 2000).

A large body of research has led to the realisation that (C)APD is not simply due to one underlying cause, but the result of a range of problems that can arise within the CANS (e.g., Berlin et al., 1972, Jerger and Jerger, 1974, Jerger and Musiek, 2000, Lynn and Gilroy, 1972, Musiek et al., 1979, ASHA, 2005, Cacace and McFarland, 2005, Campbell et al., 2011, McArthur, 2009, Musiek et al., 2010, CISG, 2012). It is characterised by poor perception of both speech and non-speech sounds (Campbell et al., 2011). At the Bruton Conference it was reported that (C)APD was considered as a deficit in processing of information that is specific to the auditory modality (Jerger and Musiek, 2000). The BSA (2011) stated that (C)APD can arise from frank lesions in the central auditory system or may co-exist with, and/or be linked to peripheral hearing loss. A variety of potential causes of (C)APD, representing a spectrum of associated disorders, prevented scientists from reaching a common consensus on evaluation and management of this intricate disorder. Rather than a spectrum, in this thesis a categorised approach is proposed as this approach may provide a
new insight into this wide-ranging disorder, as based on the main problems that this population suffer from in their routine listening environments. Hence, this new (categorisation) approach might have more potential to create a common consensus among researchers. For example, it is reasonable to divide the (C)APD population into two groups with one group having a series of problems with understanding speech in the presence of background noise as their common complaint, and those who might have problems with noise but not as a main problem (when compared to the first group). Moreover, one might categorise the (C)APD population as those with a temporal processing problem, those with a problem in their short-term memory, a problem with slow informational processing speeds, a problem with decoding of information, etcetera.

A new subcategory of CAP is spatial processing. Spatial processing is the capacity to selectively attend to an acoustic signal arriving from one location while concurrently competitive or unwanted signals, such as noise, are incoming from another location. It is based on spatial cues including ILD, ITD and spectral information of incoming signals arriving at our two ears. These signals interleave and create a complex acoustic environment. This environment has to be analysed and interpreted based on the spatial cues of individual signals, each of which need to be segregated into their original streams (Dubno et al., 2002), so they can be accurately perceived. Segregation of acoustic signals to extract the intended signal from the unintended ones is important in the processing of auditory information. Different terminology for spatial processing has been used in literature, such as spatial hearing, spatial stream segregation, and spatial release from masking (SRM).

Spatial processing disorder (SPD) is defined as a reduced capacity to selectively attend to an acoustic signal arriving from one location while concurrently interfering signals are incoming from another location (Cameron and Dillon, 2011). It reflects a deficit in SRM, hence the lack of improvement (defined as the spatial advantage) in the detection threshold when the target is spatially separated from interfering sounds compared to when it is co-located with maskers (Cameron et al., 2006, Darwin, 2008, Bronkhorst, 2000, Zurek, 1993). For example, attending to a speech signal in the presence of background noise is found to be compromised in children with SPD, and these children will experience more difficulties in listening tasks in comparison with their normally hearing peers, unless they are provided with a higher signal to noise ratio (SNR) in the classroom (Cameron and Dillon, 2008). Even children with no processing deficits have been shown to require a minimum of 15 dB SNR in
Chapter 1: Introduction

a classroom environment to obtain maximum speech recognition (Nelson et al., 2002). Given those measurements in typical school classrooms have shown SNRs ranging from 4 to -7 dB SPL (Crandall and Smaldino, 2002), it is not surprising that a low SNR, together with reverberation, has been shown to have adverse effects on academic performance and outcomes (Palmer, 1997). A person is diagnosed with SPD depending on the results of the Listening in Spatialised Noise-Sentences (LiSN-S) test. If their spatial advantage is worse than 1.96 standard deviations (SDs) from the mean for their age, they are labelled as having SPD (Cameron et al., 2014).

Spatial processing is closely dependent on selective attention. Attention is the ability of cognitive resources to process intended stimuli while ignoring other distracting stimuli (Anderson, 2004). Spatially separating target and distracting stimuli results in an improvement of the ability of the auditory system to identify the target stimulus by allocating more attentional resources to the target and fewer attentional resources to the distractors. This devotion of attentional resources to the target helps listeners to maintain their concentration on the target (Freyman et al., 2004).

Despite the large body of behavioural research into spatial processing in the context of normal-hearing population and those with a disorder in spatial processing (e.g. Ahlstrom et al., 2009, Arbogast et al., 2005, Cameron and Dillon, 2007b, Dubno et al., 2002, Dubno et al., 2008, Gelfand et al., 1988, Kim et al., 2002), little is known about the physiological mechanisms underpinning this processing. In particular, the relative contribution of different structures within the CANS to spatial processing has not yet been fully established. Moreover, to my knowledge no accepted protocols and documented procedures for recording AEPs to evaluate spatial processing, and identify populations with SPD in particular, have been established so far.

To address these issues, the aim of this thesis was to identify electrophysiological representations of spatial processing by simultaneous recording of electrophysiological activity from the brainstem and auditory cortex. In order to investigate these different electrophysiological activities, a range of commonly-used electrophysiological recordings was implemented. These recordings comprised of auditory brainstem responses (ABRs), frequency following responses (FFRs), and cortical auditory evoked potentials (CAEPs). By recording these potentials, it was expected that this project would shed light on the physiological processes mediating the ability to listen to the target signal in the presence of
distracting noise. It also was expected to identify the characteristics of electrophysiological recordings obtained from different levels within the CANS (i.e. brainstem or cortex) during the act of spatially processing an acoustic event in a normally hearing population. If physiological processes underlying spatial processing can be better identified and understood, then objective evaluation of people with SPD, particularly for evaluating very young children, may be possible.

A practical outcome of this project has been the design of a well-organised protocol for recording the brainstem and auditory cortex simultaneously to identify whether they exhibit spatial release from masking (SRM). This isochronal recording for measuring SRM required a range of methodologies, which remained challenging because of technical and physiological limitations. This simultaneous recording is quite different from Lauffer et al. (1993) who used pure tones for evoking cortical responses and filled the inter-stimulus interval (ISI) with clicks in their study. Filling the ISI with clicks resulted in an increase of cortical response latencies and a decrease of amplitude of N1 and P2 due to adaptation within CANS.

1.2. Objectives and research questions

The objective of this thesis is:

To investigate the effect of spatial separation on the detection of sounds in a competing acoustic environment, by examining the electrophysiological responses recorded from the auditory brainstem and auditory cortex using standard electrophysiological measures.

Based on this objective, the following research questions were considered:

- Do electrophysiological measures of neural responses to auditory input:
  a. Provide evidence of processing underpinning SRM in the brainstem when using speech-evoked frequency following responses?
  b. Identify processing underpinning SRM in either the brainstem or cortex (or both) simultaneously when using ton-evoked (brainstem) and cortical measures?

This objective also includes answering one minor research question:
Chapter 1: Introduction

a. Are attentional mechanisms relevant for spatial processing at the level of the brainstem?

1.3. Thesis outline

This thesis comprises of five chapters.

In this chapter, the background and reasons for undertaking this research, the objectives, and the related research questions have been outlined.

Chapter 2 presents a review of the theoretical basis of central auditory processing and its disorders by considering the neurophysiology of this process. This is followed by a review of previous research regarding spatial processing disorder as a subcategory of (C)APD, highlighting both the outcomes and limitations of previous studies. Finally, a description of electrophysiological methods that are used in this thesis is provided.

In Chapter 3, study one is presented, which investigated the adequacy of FFRs to objectively reflect spatial processing. The effects and interactions of stimulus signal-to-noise ratio (SNR), attention, and electroencephalogram (EEG) recording channels on FFR amplitudes in co-located and separated conditions is reported.

In Chapter 4, study two is presented, which investigated simultaneous recording of the brainstem and cortical potentials to investigate the physiological location of, and relative contribution to, spatial processing. Results of ABRs, FFRs, and CAEPs recorded at different stimulus SNRs, spatial and target pattern conditions are presented.

Finally, in Chapter 5, findings of this research project, discuss the results, derive our conclusions, and propose future directions for research based on our findings are summarised.
Chapter 2. Spatial processing and its disorder: Neurophysiological and electrophysiological approaches
2.1. Preface

Having laid out the experimental aims and research questions in Chapter 1, this chapter will introduce central auditory processing in general and spatial processing in particular. As this thesis will use electrophysiological signals from the auditory pathway to objectively evaluate spatial processing, an overview of the anatomy and physiology of the auditory pathway from the outer ear to the auditory cortex, relevant for spatial processing, will be provided. Then, electrophysiological measures at different levels of the auditory pathway and their capacity to represent spatial processing will be introduced. Finally, the rationale behind the use of objective responses from the brainstem and auditory cortex in this thesis will be explained.

2.2. Introduction

In many everyday situations, listeners are required to extract meaning from a degraded speech signal, often in the presence of background noise. Understanding speech in difficult listening environments is not only problematic for a person with a hearing loss, but is challenging even for normal-hearing people. Several factors contribute to this difficulty including, but not limited to, poor temporal processing, reduced audibility, insufficient segregation of competing signals and attentional overload (Baltzell and Billings, 2014, Billings et al., 2009, Hannemann et al., 2007, Mattys et al., 2009, Romei et al., 2011).

In difficult listening environments, the presence of an intact peripheral auditory pathway does not necessarily ensure adequate processing of a signal. For effective signal processing, the central auditory pathways and centres underpinning cognitive abilities of working memory and attention must also be intact and functional, as extracting vital information from a complex auditory signal relies on multi-staged processing within the peripheral and central auditory system (Bronkhorst and Plomp, 1989, Helfer and Wilber, 1990, Pichora-Fuller and Singh, 2006, Romei et al., 2011). Therefore, any abnormalities in these mechanisms and pathways can lead to communication breakdown (Schneider et al., 2010).

A person with a normal audiogram who appears to struggle in difficult listening situations, such as when there is background noise, may be suffering from a central auditory
processing disorder (C)APD. A sub-category of (C)APD, termed spatial processing disorder (SPD) has been identified by Cameron and Dillon (2008). Spatial processing is the ability to selectively attend to a target signal arriving from one source direction in the presence of distractors which are arriving from other source directions (Dubno et al., 2002).

In normal-hearing people, spatial processing provides quite a large benefit by enabling separation of the target signal from distractors. This benefit is quantifiable as an improved speech reception threshold: separation results in understanding speech more clearly than if the signal and the distractor were coming from same direction. This process is referred to in the literature as spatial processing or spatial release from masking (SRM).

Spatial processing is dependent on several factors in a listening environment. These factors include: compactness of the acoustic image, familiarity with the content of the message, visual cues, and a prior knowledge of spatial location and spatial cues (Schneider et al., 2007). Although each of these factors are important in recognising speech in complex listening environments, spatial cues play a substantial role in the processing of speech in such environments. To perform this processing efficiently, minute differences in time and intensity of incoming signals into both ears must be detected, and used to process signals accurately. Further to these binaural cues, spectral cues, which represent a direction-dependent filtering of the head, pinna and the torso- can also assist with signal processing (Blauert, 1997, Hammersho and Mo, 1996, Morimoto and Aokata, 1984). The pinna is the visible part of the external ear and has many superficial bumps and grooves (Kessel and Kardon, 1979). The torso is the central part of our body from which extend the limbs and neck (Dorland, 2011).

People with spatial processing disorder (SPD) are unable to demonstrate this spatial benefit (the extent to which the auditory system is able to separate a target signal from a distractor), despite having normal peripheral hearing (Cameron et al., 2011). It has been demonstrated that SPD can have a detrimental effect on understanding speech by normal-hearing children in the presence of background noise (Cameron and Dillon, 2008). Spatial processing disorder can be identified using behavioural assessments such as the Listening in Spatialised Noise Sentences-Test (LiSN-S) (Cameron and Dillon, 2008) and the Listening in Spatialised Noise –Tone complex (LiSN-T)(Buchholz et al., 2013). However, significant improvements in understanding and management of SPD could be achieved with the development of an objective method to detect SPD at an earlier age than is currently
possible, and that could provide detailed information regarding when or where in the auditory pathway the processing breakdown occurs. Identifying both subjective and objective measures for detecting this impaired processing could help researchers to develop training devices, and clinicians to implement and monitor the training process. Electrophysiological responses recorded in response to auditory stimuli, for example the auditory brainstem response (ABR), the frequency-following response (FFR), or the cortical auditory-evoked potential (CAEP), could potentially be used as objective tools to measure central auditory nervous system (CANS) function.

By analysing recordings made with different objective tools (e.g. ABR, FFR, and CAEP) it may be possible to evaluate where spatial processing occurs in the auditory pathway, and where any potential disorder is centred. For example, these electrophysiological measures also may help us to determine whether deficits are due to abnormal neural encoding in ascending auditory pathways, or difficulties with higher order cognitive processes needed for performing speech perception in noise (Billings et al., 2013). These tools could also provide valuable information about how large populations of neurons, recorded at the scalp, are encoding signals in noise (Billings et al., 2009, Näätänen and Picton, 1987, Vaughan and Ritter, 1970, Wolpaw and Penry, 1975).

In this chapter, I will present a brief overview of (C)APD and SPD. Secondly, a detailed review of the anatomy and physiology of the CANS will be provided, from the level of the outer ear to the level of the cortex, and in particular, the role in spatial processing of different structures from the cochlear nucleus upwards to higher order centres. Finally, the electrophysiological representation of spatial processing in general, and the possibility of recording spatial processing in the brainstem and auditory cortex specifically will be discussed, together with a rationale for choice of measure to record spatial processing in the rest of the thesis.

2.3. Central auditory processing (CAP)

Central auditory processing (CAP) refers to the perceptual processing of auditory information in the CANS and the neurobiological activities that underpin the perceptual processing. These neurobiological activities give rise to electrophysiological auditory potentials (ASHA, 2005). Auditory potentials are electrical activities of neurons arising from
acoustic stimulation (see Section 2.5) (Picton et al., 1974). CAP is responsible for a variety of auditory behaviours including sound localisation and lateralisation, auditory discrimination of speech and non-speech signals, auditory pattern recognition, the temporal aspects of audition (including, temporal ordering, temporal masking, temporal resolution, and temporal integration), and auditory performance with degraded or competitive acoustic complexes (ASHA., 1996, ASHA, 2005). Other mechanisms that are involved in acoustic signal processing are memory, learning, attention, long-term phonological representation and other neurocognitive processes on the higher level of the brain (Musiek and Chermak, 2007, Medwetsky, 2009).

2.3.1. Central auditory processing disorder ((C)APD)

Central auditory processing disorder ((C)APD) has been an extremely controversial topic in hearing science over the last fifty years (Miller, 2011). The controversy largely arises from the peculiarity of the disorder in that it is comorbid with other associated conditions (e.g. attentional deficit, learning disabilities, and language impairment), the diversity of symptoms and signs associated with the disorder, its diagnostic indication, as well as the way the disorder should be assessed and managed (Moore et al., 2013).

People who usually report listening difficulties in everyday acoustic environments, but who show basic audiometric assessments within the normal range (Hind et al., 2011), may be suffering from a (C)APD. Children and adults with (C)APD frequently have difficulty listening in noisy situations, in conversations on the telephone, following long conversations, learning a foreign language or challenging vocabulary words, memorizing spoken information (which is a deficit in auditory memory), taking notes, maintaining focus on an activity if other sounds are present, and with organisational skills, pursuing multiple directions, reading and/or spelling, processing nonverbal information (e.g., lack of music appreciation), and directed, sustained, or divided attention (Schminky and Baran, 1999).

The term “(C)APD” was first coined in 1977 at a conference in Cincinnati held in order to explore (C)APD, although the role of auditory perception in language disorders had already been investigated (Rees, 1973). Since that time, several different descriptions of (C)APD have been developed (ASHA, 2005, Jerger, 1998, Jerger and Musiek, 2000). For example, the British Society of Audiology (BSA, 2011) stated that “(C)APD has its origins in
impaired neural function and is a collection of symptoms that is usually comorbid with other neurodevelopmental disorders. It is also characterised by poor perception of both speech and non-speech sounds and usually has an impact on everyday life primarily through a reduced ability to listen, and hence to respond appropriately to sounds.

For the purpose of this thesis, (C)APD is defined according to the American Speech-Language-Hearing Association (ASHA) 2005 definition, “(C)APD refers to difficulties in the perceptual processing of auditory information in the CANS, and the neurobiological activity that underlies that processing and gives rise to the electrophysiological auditory potentials” (ASHA, 2005).

(C)APD may be diagnosed at any age of life; occur secondarily to neurological disease, insult, or injury; be of unknown etiology (Bellis and Anzalone, 2008); or co-occur with, and/or be combined with peripheral hearing loss (Alles et al., 2011). Although the main body of research and most clinical considerations have focused on children, it is well documented that (C)APD also occurs in adult and elderly populations.

Estimating the prevalence of (C)APD in different populations is still problematic due to differences in its definition. It has been approximated that the prevalence of (C)APD in school-aged children is about 2% to 5% (Bamiou et al., 2001, Chermak and Musiek, 1997), whereas prevalence within the elderly population has ranged from as low as 2% to as high as 76% (e.g., Chermak and Musiek, 1997, Golding et al., 2004, Golding et al., 2005, Stach et al., 1990, Cooper and Gates, 1991). Using 9 different sets of criteria for diagnosing (C)APD based on ASHA (2005), AAA. (2010), BSA (2011) and selected research, Wilson and Arnott (2013) reported rates of diagnosis of (C)APD ranging from 7.3% to 96.0%. This wide range depended on which criteria were applied to 150 school-aged children with normal peripheral hearing who had undergone the same (C)APD assessment.

Due to varying interpretations of how to identify the causes of (C)APD, the aetiology of this problem has not yet been fully established. However, based on accumulated data, it can reasonably be described as being attributable to abnormal neurophysiologic representation of auditory stimuli, such as inefficient inter-hemispheric transfer of auditory information, lack of appropriate hemispheric lateralisation, inaccurate synchrony of neural firing, atypical hemispheric asymmetry in neural representation of auditory signals especially speech stimuli, inaccurate temporal processing (Bellis and Wilber, 2001, Jerger et
al., 2002, Kraus et al., 1996, Moncrieff et al., 2004), and possibly, inaccurate releasing of neurotransmitters.

A pragmatic approach to identifying and treating (C)APD is to categorise different presentations of (C)APD. These categorisations should be quantifiable and measurable in the way that they would be administered in a clinical setting. However, due to the heterogeneous nature of (C)APD, different scientists and scientific societies suggest different categorisations. As an example, the BSA (2011) suggests three categories for (C)APD:

1. Developmental (C)APD: This category of (C)APD starts from childhood with normal basic audiometric results and no other recognized aetiology or risk factors. It may continue into adulthood in some cases.
2. Acquired (C)APD: This category of (C)APD starts after birth due to a known event (such as, acquired brain injury, stroke, tumours, infection, and noise induced hearing loss).
3. Secondary (C)APD: This category of (C)APD occurs due to transient or progressive peripheral hearing impairment (such as, otitis media and presbycusis).

Although they give a general insight into the disorder, the categories as enunciated by the BSA (2011) do not provide an operational basis for clinical usage. For example, they are unable to explain whether there is any difference between the types of (C)APD that each category represents. Theoretically, based on different approaches, one can assume different etiological categorisations of (C)APD. However, by considering (C)APD as a common problem, it potentially may be possible to create operational categories. A common complaint of people with (C)APD is the difficulty in understanding speech where background noise is present. For example, children with (C)APD demonstrate difficulties in understanding speech when confronted by a noisy environment, for example when they are at school (Chermak et al., 1999, Chermak et al., 2002).

Recent studies in some children identified with (C)APD have shown that it may result from a deficit in the mechanisms that are responsible for SRM (Cameron and Dillon, 2007b, Cameron and Dillon, 2008, Jerger, 1998). These children are unable to differentiate target from non-target sounds in complex acoustic environments. They experience difficulties when concentrating on target sounds and ignoring distractor noises. The condition
specifically has been known in the literature as spatial processing disorder (SPD). This condition results in children mixing up both targets and distractors because they have problems with the spatial processing of complex acoustic environments. This research is in line with other current research that will result in the emergence of a new operational category for (C)APD, known as SPD. In the following section, the fundamental concept of (normal) spatial processing will be discussed.

2.3.1.1. Spatial processing

Spatial processing refers to a series of processing performed by the auditory system on a target sound to determine the location of that target sound in space. The location of the target in space usually varies relative to other interfering sound sources which may occur very close to the target sound or may arrive from different directions to that of the target sound (Bregman, 1990, Moor, 1997). The target and distracting sounds that arrive at each ear will be a single acoustic complex that results from the addition of the sounds from multiple sources and their reflections. With increasing complexity of the acoustic environment, the ability of the auditory system to extract the target sound reduces proportionally. In our everyday lives, most of us are routinely confronted with this type of environment. The main reasons why we cannot easily understand the target sounds are the masking effects of distracting sounds, the inability to separate target from distracting sounds, switching attention back and forth between two target and distracting sound streams, and the similarity of target and distracting sounds (Schneider et al., 2007).

Two types of cues are critical in spatial processing, the binaural and monaural cues (Yost and Gourevitch, 1987). The main binaural cues are interaural level differences (ILDs) and interaural time differences (ITDs) which are disparities in intensity and time of impinging sounds on the ears, respectively (Middlebrooks and Green, 1991). The information in low-frequency regions below 1-2 kHz (Strutt, 1907) and high frequency superimposed on low frequency carriers (Van De Par and Kohlrausch, 1997) in the azimuth plane is conveyed by ITD cues. Conversely, for high-frequency sounds above 1-2 kHz (Strutt, 1907) and low frequency for near sources (Shinn-Cunningham, 2000), the auditory system computes the ILD cues in the azimuth plane. The main monaural cues that contribute to spatial processing are the monaural spectral information (Wightman and Kistler, 1997),
monaural distance cue (Zahorik et al., 2005) which is the overall received sound level (Warren, 1999) and, in reverberant environments, the direct-to-reverberant ration (DRR) (Larsen et al., 2008). While monaural cues play a more important role for people with unilateral hearing loss, in normal-hearing individuals these cues provide redundant and complementary information to binaural cues, enhancing understanding in noise.

Spatial processing relies on selective attention. Selective attention, which is an executive, top-down, mechanism, is defined as the ability of the cognitive mechanisms to focus on sensory information that is pertinent to an intended stimulus (Gazzaley and Nobre, 2012). This sensory information is derived by bottom-up mechanisms via a dynamic relationship with the top-down mechanisms. In the auditory modality, selective attention to sound refers to a listener’s ability to attend to a target sound and ignore the distracting sounds at different locations in an acoustic environment. The ability to identify an intended acoustic signal in the presence of distractors is based on different combinations of auditory information, such as spatial location cues and other relevant information of the intended signal, which convey by the auditory pathways towards the higher processing centres. It is well documented that the auditory system has neural networks, known as efferent pathways, which convey feedback information from higher structures to the lower structures. These neurons enervate to neural populations at the levels of the ascending auditory system including the thalamus, the inferior colliculus (IC), the superior olivary complex (SOC), and the cochlear nucleus (CN) (Diamond et al., 1969, Mulders and Robertson, 2000, Weedman and Ryugo, 1996, Winer et al., 1998). The efferent neurons influence many characteristics of neural responses such as filtering (Diamond et al., 1992) and increasing their tuning (Villa et al., 1991) at the subcortical levels.

Traditionally, two spatial conditions exist; co-located and separated. Experimentally, a researcher usually presents the co-located target and distracting sounds from the front (0 degrees). In the separate condition, the target is usually again presented from the front, but the competing sounds from some other angle(s), often ±90° azimuth. In addition to spatial separation, there are other factors that can help in spatial processing or SRM, including informational masking (usually masking of speech, by speech), acoustic image compactness, familiarity with the content of the message, prior knowledge of spatial location, and visual speech cues (Schneider et al., 2007).
Binaural processing enables a listener to detect the target sound by using spatial cues to differentiate the target from the distractor in complex acoustic environments. In this condition when the SNR in the critical band (based on the filter bank theory) is low and the energy of distractors arriving at the ears is the dominating factor, the activity pattern of the neurons that detect ITD cues of the signal will also detect distracting ITD cues, reducing the salience of the inter-aural cues for the target sound. One hypothesis to explain how a complex acoustic environment is processed spatially is that the binaural mechanism operates separately in each frequency channel for which ITD differences between the target and masker exist (Akeroyd, 2004, Edmonds and Culling, 2005).

Spatial separation of target and distractor improves the detectability of the target and/or intelligibility of speech stimuli. To demonstrate this improvement in the detectability of target sounds, Freyman et al. (1999) and (2000) in an experiment presented in a first condition the target and distractors from the front and presented in a second condition the target from the front and distractors both from the front and from the right side of the participants. In the latter condition, the distractor presented from the right preceded in time the distractor from the front. This temporal leading of the distractor resulted in a shift of the perceived distractor away from the target and hence resulted in better understanding of the target (Freyman et al., 2000, Freyman et al., 1999). This contrast is known as spatial advantage. However, if the nature of the distractor and target were dissimilar, for example, a speech target in broadband noise, this dislocation had little effect on perception. The greater the similarity of target and distractor, the greater the effect of the distractor and the larger the amount of spatial advantage, showing the important role of spatial separation in providing release from informational masking (Brungart and Simpson, 2002, Culling et al., 2004, Durlach et al., 2003a, Durlach et al., 2003b, Freyman et al., 2001).

While the role of fundamental cues in sound localisation in simple environments is well established, little is known about complex acoustic environments, such as environments with high levels of room reverberation. In these scenarios, the listener receives the direct target sound at the ears convolved with several copies of the original sound source resulting from echoes rebounding from the wall surfaces. Shinn-Cunningham et al. (2005b) found that reverberation alters, independently at each frequency, sound magnitude and phase, monaural spectrum cues, ITDs and ILDs. These effects are based on
the source location relative to the listeners and on their position in the room as well. Consequently, binaural spatial processing and spectral cues are essential when there are several simultaneous sources. In these scenarios, the capability of the auditory system is quite important to process target (speech) sounds (Best et al., 2008, Brungart and Simpson, 2007) and to determine source localisation (Drullman and Bronkhorst, 2000, Hawley et al., 1999) precisely.

Several studies have suggested that differences in some physical aspects of the stimuli can help in understanding the target such as: talker gender, sound level, voice timbre, disparity in fundamental frequency, and perceived location (Bird and Darwin, 1997, Brungart et al., 2001, Darwin et al., 2003, Darwin and Hukin, 2000, Drennan et al., 2003, Durlach et al., 2003b, Kidd et al., 1994). Apart from the individual characteristics of talker and physical characteristics of target signal, there are some other factors that can also help in understanding the target signal in the presence of distractors, particularly those distractors that interfere with the content of the target. These factors are image compactness (i.e., a compact spatial image of the target in the presence of a diffuse spatial image of the distractors), previous knowledge of the content of the message, previous knowledge of target speech location, and availability of visual aspects of target speech (Schneider et al., 2007).

During the spatial processing of complex acoustic elements of speech in the presence of noise, different regions of the auditory pathways behave in different ways. Temporal processing of the transient part of the consonant element is most strongly represented and might be further improved at the cortex level (Heil and Irvine, 1997). This mechanism is implemented by neural inhibition and the conjunction of multiple inputs possibly contributing to representation of spectro-temporal aspects of consonants through the voicing and place of articulation cues (Steinschneider et al., 1982). On the other hand, Cunningham et al. (2002) revealed that the steady-state response of the vowels showed greater resistance in the presence of noise at subcortical regions compared to transient responses. This resistance is partly because the neural discharge of most units become saturated in noise and therefore in order to preserve the neural representation of the vowel, the auditory system extracts the temporal code from the steady-state portion of the vowel by phase-locking of neural activity through the entire duration of the vowel (Delgutte and Kiang, 1984).
In addition to temporal features of the vowel stimulus, the synchronisation of neural population firing to represent the harmonics around each formant frequency is well established (Shamma, 1985). In another attempt to identify the role of neural fibres in sound localisation at different levels of the CANS, Moore et al. (1974) examined the role of commissural pathways of the auditory system in cats. Their results demonstrated that the superior olivary complex (SOC, please refer to Section 2.4.2.2 for more detail) plays a critical role in encoding cues of sound localisation and the most important crossed pathway that makes such interaction feasible in the level of the medulla.

In summary, spatial processing is one of the most complex activities and functions of the CANS, relying on many different cues, most critically spatial cues, to enable a person to understand speech in background noise.

2.4. The auditory system and its relevance for spatial processing

To record the auditory evoked potentials in this project, all stimuli were delivered under headphones to the ear canal, and via the peripheral auditory system toward the central auditory system. To provide a basis for our discussion of the genesis of spatial processing in the auditory system, it is appropriate to present a brief review of the structures and functions of the auditory system, with a focus on the relevance of each part for spatial processing.

The human auditory system consists of two main parts, the peripheral auditory system and the central auditory system, both of which will be described in the following sections.

2.4.1. The peripheral auditory system

Figure 1 illustrates different parts of the auditory periphery. This portion of the auditory system is usually subdivided into the outer ear, the middle ear, the inner ear and the auditory nerve.
Chapter 2: Spatial processing and its disorder

2.4.1.1. Outer ear

The outer ear consists of three components: the auricle (a protruding fold of cartilage-supported skin on each side of the head); the concha (the shell or funnel-shaped part); and the external auditory meatus (a tube like structure that is slightly “s” shaped in humans) (Figure 1).

![Figure 1. The structure of the peripheral human ear. The pinna, external auditory canal, tympanic membrane, ossicular chain, and the inner ear (Kessel and Kardon, 1979).](image)

The primary function of the outer ear is to pick up sound and funnel it along the ear canal towards the tympanic membrane. The head and external structures of the outer ear change the amplitude and phase of some spectral components of the sound. This spectrotemporal characteristic is known as the head-related transfer function (HRTF) (Figure 2). The HRTF for sounds from any angle and distance describes the sound pressure at the eardrum divided by the sound pressure in the undisturbed sound field. It is most important for high-frequency sounds, as their wavelengths are close to the size of head, pinna and torso (Blauert, 1969, Blauert, 1997, Klumpp and Eady, 1956).

2.4.1.2. Middle ear

The middle ear is an air-filled cavity situated between the outer and inner ear. It is connected to the pharynx via the Eustachian tube, which plays a role in the equalisation of pressure between the external canal and middle ear. Its distal wall comprises the tympanic
membrane, which separates it from the external auditory meatus. The ossicles, including the malleus, the incus, and the stapes, are situated within the middle ear. The head of the malleus is in contact with the tympanic membrane, while the footplate of the stapes is situated in the oval window, the entry to the cochlea or inner ear.

Figure 2. The head-related transfer function (HRTF) shows the amplitude spectra at each ear (measured within the ear canal) on an arbitrary decibel scale of the dummy head at different distances. The sound at the left ear is attenuated relative to the right ear, especially at higher frequencies (Potisk, 2015).

Acoustical energy arrives at the tympanic membrane (TM); a thin, cone-shaped three-layer membrane. Vibration of the TM is then transmitted through the middle ear by means of the ossicular chain. The ossicular chain acts as a first class lever, converting the vibration of the TM to mechanical energy that is then transmitted to the fluid-filled cochlea by means of vibrations of the footplate of the stapes located against the oval window membrane. The area ratio between the TM and oval window membrane, together with the lever advantage provided by the ossicular chain provides the necessary impedance coupling for effective transmission of the acoustic energy between the air-filled ear canal and the denser fluid-filled inner ear (Sohmer et al., 2000, Stenfelt and Goode, 2005).

A normally functioning middle ear plays a critical role in transmission to the CANS of an accurate representation of the acoustic energy arriving at the auditory periphery. The presence of middle ear pathologies, for example acute or chronic otitis media, or a cholesteatoma, can change the conductivity characteristics of this acousto-mechanical transducer. These pathologies may result in a failure to transmit important spectral or temporal cues necessary for processing of complex acoustic environments. Otitis media is the pathology most likely to result in a spatial processing disorder (SPD) in children. It is
reported that 50% of the children diagnosed with SPD in different studies at the National Acoustic Laboratories also had a history of chronic otitis media (COM) (Cameron and Dillon, 2013, Cameron et al., 2014).

2.4.1.3. Inner ear

The inner ear consists of three parts: the semicircular canals; the vestibule; and the cochlea (Figure 3). The semicircular canals and the vestibule are primarily associated with the sensation of balance, and as such, are beyond the scope of this thesis.

The cochlea, which contains the sensory organ of hearing, is divided into three compartments (Figure 3): scala vestibule, scala tympani, and scala media, or as sometimes called the cochlear duct. Within the scala media lays the organ of Corti, bounded above by Reissner’s membrane, bounded below by the basilar membrane. On its lateral wall, the stria vascularis can be found. The scala media also houses outer hair cells (OHCs) and inner hair cells (IHCs). The OHCs and IHCs, with geometrical tiny cilia protruding from their surface, act as a biomechanical amplifier and transducer, respectively (Johnstone et al., 1986).

The basilar membrane (BM), which functions as a tonotopically organized spectral analyser, has shorter, stiffer transverse fibres at the base of the cochlea, and longer and
more flaccid fibres at the apex of cochlea. This heterogeneity in texture results in the BM vibrating at different resonant frequencies along its length (Oghalai, 2004).

The cochlea has two primary functions: frequency analyses and mechano-electrical energy conversion. During frequency analysis of incoming acoustic stimuli, different acoustic frequencies result in maximal displacement of the basilar membrane in different regions tonotopically related to frequency. The upwards vibration of the basilar membrane results in upwards movement of the hair cells, and deformation of the stereocilia against the tectonic membrane. This movement in turn results in the flow of potassium ions, and hence changes to the permeability of the cell membrane and a potassium/sodium exchange in the stimulated region. Changing the ions inside OHC changes the receptor potential of the cells, which drives oscillations in the cell’s length, with the same frequency as the dominant sound at that point in the cochlea, and so provides mechanical amplification, enabling easier detection by the IHCs. Increasing the potential of IHCs, on the other hand, then triggers neurotransmitters to be released which generates an action potential in the auditory nerve (Davis, 1983, Gold, 1948, Brownell et al., 1985, Müller, 2008).

An intact cochlea enables relatively accurate translation of the spectral and temporal features of the acoustic signal into neural signals via stimulation of the hair cells. The resonant properties of the basilar membrane changes along its length, and as a result, the basilar membrane is tonotopically organised. Cochlear pathology, on the other hand, may affect either the IHCs or the OHCs. Typically, damage to the OHC’s leads to a loss of frequency selectivity and sensitivity to sound, whereas damage to the IHCs results in a reduced ability to encode the particular frequency resonant in the area of the cochlea damage. Damage in the cochlea impairs the ability of the cochlea to encode spectral elements of the sound stimuli, and as such affects the auditory pathway’s ability to undertake a spectral decomposition of the sound. Lack of normal spectral decomposition of the sound results in a blurred and distorted mapping of the sound energy onto the auditory nerve array (Musiek and Chermak, 2007).

2.4.1.4. Auditory nerve (AN)

The innervation of the cochlea consists of two types of auditory nerve fibres, afferent and efferent fibres. Afferent fibres transmit peripheral information to the central
auditory system. Efferent fibres transmit information from higher neural centres to exert some descending control on the functioning of the cochlea.

Afferent fibres include radial (type I) and outer spiral fibres (type II). The type I fibres comprise about 85% to 95% of the afferent fibres and exclusively innervate IHCs at the base with a several-to-one arrangement. The type II fibres, about 5% to 15% of remaining afferent fibres, innervate the basal area of the OHCs as one-to-several connections (Spoendlin, 1969).

The efferent fibres are a series of nerves (the olivocochlear bundle, OCB, as shown in Figure 4) that descend from the superior olivary complex (the SOC, to be introduced in Section 2.4.2.2) towards the cochlea. The OCB synapses on afferent fibres where they are leaving the IHC and directly on the basolateral surfaces of the OHC (Iurato et al., 1978, Smith, 1961). The effects of the OCB on the afferent fibres that innervate the IHCs can be excitatory or inhibitory (Eybalin, 1993). The OCB has an inhibitory effect on OHC’s motility through hyperpolarisation, which reduces the gain of the cochlear amplifier (Guinan, 1996), resulting in reduced excitation of the IHC (Elgoyhen et al., 1994).

Several investigators have examined the effects of background noise on the dynamic ranges of auditory nerve fibres in different animal studies. In general, they have found that the dynamic ranges of most fibres shifted to higher average levels in the presence of wideband noise. Changes in the dynamic range of fibres can negatively influence the sound intensity coding at low levels, although it can improve it at higher levels (Costalupes et al., 1984, Geisler and Sinex, 1980, Gibson et al., 1985).

In animal studies, it has been shown that the OCB can enhance the auditory nerve responses to transient sounds in noisy backgrounds (Dolan and Nuttall, 1988, Kawase et al., 1993, Winslow and Sachs, 1987). OCB stimulation inhibits the auditory nerve response to the background noise, which results in restoring the auditory nerve fibre’s dynamic range to some extent. This phenomenon is known as the OCB unmasking effect (Guinan, 1996). It is highly probable that OCB could play a role in suppressing and adapting of energetic masking (EM) which is a masking that results from interaction between target and distractors at the peripheral hearing system. Based on the suppressive and adaptive masking effect, OCB by increasing the overall target signal to distractors ratio facilitates better detection of the target signal hence improves selective listening (Kawase et al., 1993, Kawase and Liberman, 1993, Kawase and Takasaka, 1995).
2.4.2. The central auditory system

The central auditory system comprises of the auditory pathways and the higher brain structures. The auditory pathways will be briefly outlined in this section and are composed of the cochlear nucleus (CN), the superior olivary complex (SOC), the lateral lemniscus (LL), the inferior colliculus (IC) and the medial geniculate body (MGB). The auditory cortex (AC) also will be discussed.

2.4.2.1. Cochlear nucleus (CN)

As the auditory nerve fibres enter the cochlear nucleus (CN), they project onto multiple target nuclei: the anteroventral cochlear nucleus (AVCN), the posteroventral cochlear nucleus (PCN) and the dorsal cochlear nucleus (DCN). Each axon of the auditory nerve has different branches that innervate all the nuclei and so constitute parallel processing (Rhode, 1991).

The cochlear nucleus (on each of the two sides of the auditory pathway) is composed of several types of neurons that are designated by their dendritic configurations. At least
twenty-two different types of neurons have been anatomically identified (Brawer et al., 1974). Of these, there are four types of principal cells in the CN:

- **Bushy cells (spherical and globular) found in the AVCN** (Cant, 1991),
- **Stellate cells found in the AVCN** (Trussell, 2002),
- Octopus cells found in the PVCN (Rhode, 1991); and
- **Fusiform or pyramidal cells** in the DCN (Young and Brownell, 1976).

Discussion on function of these neural cells is beyond the scope of this thesis.

The cochlear nuclei are the first relay station in the auditory system, and receive information only from the ipsilateral auditory nerve. Based on their anatomical shapes, physiological response type, response area characteristics, and particular synaptic connections, each neuron of the cochlear nuclei has a specific structural and functional characteristic. These distinctive features enable them to process and convey different aspects of auditory information with high fidelity to the SOC (Section 2.4.2.2) and the IC (Section 2.4.2.4) (Wickesberg and Oertel, 1988, Wickesberg and Oertel, 1990, Young and Brownell, 1976, Young and Sachs, 2008, Young and Voigt, 1982).

As the preliminary stage in spatial processing, the CN plays a critical role in providing initial information of spatial cues to the higher regions. Having an accurate encoding of ITD and ILD cues is dependent on how accurately the CN processes target signals and distracting noise.

Cochlear nuclei neurons can enhance the SNR relative to the auditory nerve in the presence of background noise. This enhancement of the SNR occurs by decreasing the average firing rate of both signal and noise and increasing synchronous response to the signal (Frisina et al., 1997). It seems likely that dendritic filtering and postsynaptic membrane characteristics of the DCN fibres, augment synchronous responses to the signal, whereas interneuron inhibitory inputs may decrease the average responses to signal and noise (Frisina et al., 1994). In contrast, Masterton et al. (1994) in an animal experiment reported that the DCN was not specialised for intensifying the detection of acoustic signals in the presence of background noise.

In an attempt to identify whether DCN damage can influence sound localisation, Sutherland et al. (1998) did not report any deficit in localisation following DCN lesions in cats. May (2000) evaluated this issue in an animal study, and concluded that processing spectral cues in the DCN is critical for determining the elevation of sound sources, and that
localisation of a stimulus in the azimuth plane may depend on ILD and ITD information conveyed to binaural neurons in the SOC. Human studies have shown that the spectral and binaural information convey different cues in localisation of auditory signals. When HRTFs are low-pass filtered, humans are able to accurately determine the sound source in azimuth plane, but show large mistakes in identifying stimuli in the vertical plane (Middlebrooks, 1992). One way to differentiate between the role of spectral and binaural information in elevation and azimuth planes is to assess a subject monaurally, by putting a plug in one ear. In May’s study, cats maintained good orientation precision in the hemifield of the unblocked ear. However, a lesion of the DCN resulted in failure of the cats to orient towards a sound source in the vertical plane, but preserved their localization of sound in the azimuth plane.

In general, it is thought that the DCN is responsible for vertical localisation and for enhancing signals in noise, while the PVCN is responsible for temporal processing, and the AVCN is responsible for phase-locking and providing information for binaural mechanisms to compute horizontal sound source location (Rhode, 1991, Rhode et al., 1983, Wickesberg, 1996, Wickesberg and Oertel, 1988, Wickesberg and Oertel, 1990, Young and Brownell, 1976).

2.4.2.2. Superior olivary complex (SOC)

The superior olivary complex (SOC) is the first relay station of the auditory pathways that receives information from both ears, and as such, enables comparison of binaural information. The SOC includes four nuclei: the lateral superior olive (LSO); the medial superior olive (MSO); the trapezoid body (TB); and the preolivary nuclei. Considering the scope of this thesis, the role of the MSO and LSO in spatial processing will be focused.

2.4.2.2.1. Medial superior olive (MSO)

The medial superior olive (MSO) receives direct connections from the AVCN fibres on both sides of the auditory pathway. Both contralateral and ipsilateral fibres of the AVCN (which converge on the synapse of the MSO cells) are excitatory, so their inputs add together to create a stronger firing pattern. Neural tunings in the MSO are mostly biased toward low frequencies (Guinan et al., 1972b) and are assumed to encode ITDs (Goldberg
and Brown, 1969, Yin and Chan, 1990). In humans, this nucleus is larger than the LSO due to the critical roles of low frequency signals in hearing (Bazwinsky et al., 2003).

The MSO neuron in humans can respond to interaural time delays of about 10-10 μs and hence enable a sound source to be located by the listener with an accuracy of a few degrees (Klumpp and Eady, 1956). This sensitivity could be explained by a coincidence detectors model which is proposed by Jeffress (1948) (Figure 5). In this model, each neuron has an inherent delay achieved by variations in its axonal length, aligned with the opposite ear to counterbalance the acoustic delays created between two ears from a sound source in the horizontal plane. Based on Jeffress’s model, if these inputs to the MSO units are in phase, the action potentials optimally sum together, and evoke maximum action potential in the units, whereas if they arrive at different times no action potentials are created (Kuwada and Yin, 2012).

2.4.2.2.2. Lateral superior olive (LSO)

The lateral superior olive (LSO) receives both direct and indirect, via the MNTD, connections from the VCN fibres bilaterally. Most of its ipsilateral inputs are excitatory (Boudreau and Tsuchitani, 1970, Glendenning et al., 1985, Shneiderman and Henkel, 1985, Smith et al., 1993) and most of its contralateral inputs are inhibitory (Boudreau and Tsuchitani, 1970, Glendenning et al., 1985, Smith et al., 1998, Tolbert et al., 1982). Its neural tunings are mostly biased toward high frequencies (Guinan et al., 1972b) and it is thought to encode ILDs (Boudreau and Tsuchitani, 1968). In humans, this nucleus is smaller than in some animals like mice and rats, which have hearing biased towards high frequency signals and thus have a larger LSO (Bazwinsky et al., 2003).
In an animal experiment, trying to identify whether single LSO cells respond to change in sound source azimuth in a manner expected of a binaural inhibitory-excitatory cell, (Tollin and Yin, 2002b) found that LSO units in cats are sensitive to ILDs, and that contralateral inhibition shapes spatial receptive fields in the azimuth. In part of Tollin and Yin experiment, using 16 kHz tones with a duration of 300 ms, by keeping the excitatory ipsilateral stimulus constant, the inhibitory effect of the contralateral input increases as its intensity level increases. In another study, using 200 ms broadband noise, and taking advantage of a virtual space technique (convolution with a HRTF), Tollin and Yin (2002a) demonstrated that for sounds located in the ipsilateral hemifield, the LSO units responded for the entire duration of the stimuli with a robust discharge and as the source moved towards the midline the responses decreased. On the other hand, for sounds presented from sources contralateral to the midline under the virtual condition, there were no detectable responses. Nevertheless, occurring shortly after 200 ms, there were clear offset responses due to removing the inhibitory effect of contralateral units. They concluded that these behaviours are indicative of the inhibitory-excitatory (IE) nature of the examined cell (Figure 6). In fact, the units respond to stimuli presented at virtually every azimuth, but they are more sensitive to the stimuli presented in the ipsilateral hemifield.

In this thesis, real acoustic environments will be simulated by presenting target and distractor stimuli with spatial cues that correspond to different azimuths. Two spatial conditions; co-located and separated will be used. Based on Figure, in the co-located condition, it will be introduced the target and distractors from 0° azimuth angle, and in the
Chapter 2: Spatial processing and its disorder

separated condition, it will be introduced the target and distractors from 0˚ and ±90˚ azimuth angles respectively.

Both EE units and IE units might contribute to a more robust response in the separated condition than in the co-located conditions. Firstly, in the co-located condition, both the target and the masker will stimulate EE units, so the portion of the response phase-locked to the target will likely be weaker than in the separated condition. Secondly, in the separated condition, during the brief moments when one ear has a better SNR than the other ear, IE units that preferentially respond to input from one side will respond to the target in a way that is not possible in the co-located condition, where both ears have the same SNRs at all times.

![Diagram of responses of LSO cell to 200 ms broadband noise introduced from five different azimuths in binaural (A, normal) and monaural ipsilateral ear only (B, ipsi-only) conditions.](image)

**Figure 6.** Responses of LSO cell to 200 ms broadband noise introduced from five different azimuths in binaural (A, normal) and monaural ipsilateral ear only (B, ipsi-only) conditions. Negative azimuths indicate sound sources in the ipsilateral sound field. The stimulus was presented at about 20 dB above threshold at (0, 0) in the ipsi-only condition and these graphs demonstrate 20 presentations of the stimuli at each azimuth (Tollin and Yin, 2002a).

After initial encoding of spatial cues in the MSO and the LSO, the incoming signal proceeds toward the next relay station, the lateral lemniscus (LL), which will be reviewed in the next section.
2.4.2.3. Lateral lemniscus (LL)

The lateral lemniscus (LL) is made up of different axonal projections from the cochlear nuclei and the SOC (Schwartz, 1992). Cell bodies of the LL units divide into two main regions, dorsal and ventral nuclei of lateral lemniscus (DNLL and VNLL, respectively). Cell types found in the VNLL and DNLL include stellate, multipolar, globular, spherical, and elongated cells (Covey and Casseday, 1986).

The VNLL exhibits various types of nerves in different species. In mammals, the ventral part contains spherical bushy cells similar to the VCN (Schofield and Cant, 1997). This tract receives most of its excitatory contralateral inputs from the AVCN and partly from the PVCN via the trapezoid body (TB) (Glendenning et al., 1981, Schwartz, 1992) and may also receive some fibres from various segments of the contralateral SOC (Buser and Imbert, 1992). This kind of network connection gives rise to the VNLL, as a monaural nucleus with specialized synapses for coding the temporal and spectral characteristics of sound coming from the opposite side. In recordings from the VNLL in cats and bats, two patterns of responses to tones have been described. Some neurons respond to tones with regular, sustained firing and are sharply tuned, while others respond with a sharply timed action potential at the onset of a tone and are broadly tuned (Aitkin et al., 1970, Covey and Casseday, 1991).

The DNLL receives more inputs from lower structures and the contralateral LL nuclei than the VNLL. Primary excitatory inputs are received from the contralateral VCN and the contralateral LSO. This tract also receives inhibitory inputs from the ipsilateral LSO and the contralateral DNLL via the Commissure of Probst. The commissural connection between the two DNLL is a substantial neural connection. Many of the ascending afferent fibres within the auditory system below the LL terminate, and/or pass through this tract while others bypass the DNLL and VNLL (Glendenning et al., 1981, Schwartz, 1992) (Figure 7).

Glendenning et al. (1981) reported that most of the DNLL units respond to binaural intensity disparities. However, compared to the LSO units, which receive contralateral inhibition and ipsilateral excitation, they exhibit excitatory and inhibitory responses to intense sounds coming from contralateral and ipsilateral pathways, respectively.

Because of different types of neurons and a variety of performance characteristics, including transient and sustained sound processing activity, sharp and broad frequency
tuning, and monaural and binaural sensitivities (Aitkin et al., 1970), it seems likely that the LL is involved in sound localisation, spatial processing, and temporal and spectral encoding. The VNLL units receive input from the contralateral PVCN (octopus cells) (Schofield and Cant, 1997) and perform temporal coding (Covey and Casseday, 1991). They have multipolar cells which respond to tones with regular, sustained firing (Aitkin et al., 1970) and have highly organised discharge rates which, although they are initiated by an input signal, have firing rates that are controlled by the electrical properties of their membranes rather than by periodicities in the input signal (Oertel, 1985). Consequently, they might be specialised for encoding spectral cues, which are important in sound localisation of elevation and whether the source is located in front or back of the listener. In contrast, the DNLL units receive their inputs binaurally and help with processing ILD and ITD cues.

The LL units are not directly involved in the generation of a new code. However, they play a vital role for the IC ipsi- and contralaterally to generate not only new spatial codes for the determination of the sound source but, more importantly, to develop specific coding for the representation and direction of the motion of the sound source. These dynamic codes

---

Figure 7. Different connections of the LL. It receives different projections from the SOC and some of the projections from the cochlear nucleus. These connections will be relayed to the IC (Dafny and Amini, 2011).
are critical for the identity of the target signal in a multisource listening environment in which the relative location of target and non-target interchange dynamically. Pickles (2008) reported that the crossed inhibitory input from the DNLL on the opposite side to the IC increases the neural contrast between the responses to sound sources localised in different locations in the azimuth plane.

To sum up, the LL is a part of the auditory pathway located between the SOC and the IC. It has two nuclei, the VNLL and DNLL. The VNLL conveys spectral cues and DNLL conveys binaural cues while exerting some refinement processing on them. This relay station contributes to encoding of motion and direction of motion of the sound sources in the IC units, which will be discussed in the following section.

### 2.4.2.4. Inferior colliculus (IC)

The inferior colliculus (IC) is the primary centre for convergence of auditory information relevant to identification and localisation of a sound source. It also responds to temporally complex stimuli. It is assumed that this region is a transitional centre where the response changes from being dominated by simple acoustic characteristics to starting the integration of acoustic features to compose an auditory object. The IC receives modulatory inputs from the auditory cortex (AC, see Section 2.4.2.6), which is relayed down to the main centres that the IC receives inputs from. This creates a bilateral circuit that helps the IC neural structures to transform and enhance the sensory representations before delivering them to the higher structures.

The IC is one of the two colliculi, the other one being the superior colliculus, which is one of the main centres of reflex integration with inputs from the visual system. The IC has three main parts: the central nucleus (ICC), the external cortex (ICX), and the dorsal cortex (ICD) (Morest and Oliver, 1984). These parts each have a different pattern of synaptic connections, different cell architectures, and hence different functions. The central nucleus is known as the lemniscal nucleus because it receives a large amount of auditory input from the LL. The other two nuclei are known as extralemniscal nuclei and transit both auditory and non-auditory information, hence they are known as multisensory auditory pathways (Moore, 1987).
There are two main input projections to the IC from lower structures, including excitatory and inhibitory units. The major excitatory inputs come from the contralateral CN, the contralateral LSO, the ipsilateral MSO, and the ipsilateral VNLL. Inhibitory inputs are projected from the DNLL, the ipsilateral LSO and the ipsilateral VNLL (Merchan et al., 2005, Oliver et al., 1994, Winer et al., 1996). There is also a projection from the contralateral IC (Pickles, 2008).

The IC uses common terminal fields to receive inputs from different regions and converges them as one output stream. Schofield (2005) reported that in the IC, the terminal fields of the ITD-detector MSO and that of the ILD-detector, LSO are shared in some regions in the IC and divided up anatomically in other regions in the IC. It has also been found that the binaural and monaural information combine at the level of single neural cells in the IC (Ahissar et al., 1992, Oliver and Huerta, 1992, Vater et al., 1995).

The IC units receive information regarding sound source location from all four principle nuclei for spatial processing. These nuclei comprise the DCN, which responds to sound location, exploiting monaural spectral cues; the MSO, encoding ITD cues; the LSO, encoding ILD cues; and the DNLL, which responds to both ILD and ITD cues. In the low-frequency region, which commonly represents ITD cues, there are projections from both the MSO and the LSO on the same side. In the region where neurons are commonly tuned to high-frequency stimuli, representing ILD cues, there are neural inputs from the DCN and LSO both contralaterally and ipsilaterally (Loftus et al., 2004).

Encoding of movement direction of sound sources is another function of the IC, which is performed by binaurally-sensitive neural cells (McAlpine and Palmer, 2002, Spitzer and Semple, 1993, Wagner and Takahashi, 1992, Wilson and O'Neill, 1998). This encoding is achieved by receiving neural signals representative of past acoustic input, as well as of the current acoustic input, encoding at different rates and representing the direction of sound source displacement. In addition, some units in the IC are sensitive to dynamic aspects of changes in binaural phase. Such dynamic aspects are not represented in the SOC units.

Another line of spatial processing in the IC is dependent on binaural cues. The binaural characteristics of most common IC units (EI) reveal that they are inhibited by ipsilateral inputs from the MSO and DNLL, and excited by contralateral inputs from the LSO (remember that the relationship is reversed from IE at the level of LSO to EI at the level of IC). Some units (EO) are driven only by monaural stimulation, directly from the contralateral
CN, and from the ipsilateral VNLL. Finally, the remaining units in the IC (EE) are excited by sound coming from both ears via the MSO (Irvine and Gago, 1990).

It has been explained in the preceding sections that the fundamental processing of binaural cues takes place before the IC anatomical regions. The possible incorporation of multiple binaural junctions together with monaural inputs raises the question of what this amalgamation accomplishes with respect to sound localisation at the level of the IC. Several ipsilateral and contralateral inhibition inputs project onto the IC neurons. The activity of the binaural and monaural units are shaped by these inhibitions and result in refining the neural representations of sound location, and the temporal and spectral properties of a signal (Covey et al., 1996, Grothe et al., 1996, Melssen and Epping, 1992, Van Stokkum and Melssen, 1991). Localisation of a sound source is quite straightforward, relying on ILD or ITD cues (Fuzessery et al., 1985, Wenstrup et al., 1988b, Wenstrup et al., 1988a).

Even in complex environments, individual neurons integrate ILD cues from a wide range of frequencies to create a code for location. Takahashi and Keller (1994), using multiple sound sources, have studied ITD cues in low frequencies for a single IC neuron’s receptive field in the barn owl. Where identical spectral and temporal patterns were presented to the two ears, the ITD was summative and the response to the two sounds represented a single phantom source. The introduction of differences in the spectral or temporal pattern of one of the sources resulted in separated responses to each source and the phantom sound disappeared. Furthermore, the neuron’s receptive fields are also affected by the temporal factors of the preceding sound (Yin, 1994) or by the movement of the sound source (Wagner and Takahashi, 1992). Thus, the neuronal response to a sound source originating from a specific location is influenced by several factors including amplitude, stimulus type, and other sound sources that are separated in either the time or space domain (Grothe et al., 1996).

In another study, Grothe et al. (1996) computed ILD with a combination of ongoing noise in different locations and a pulsed signal. They found that the presence of the noise affected the receptive fields of all types of neurons, regardless of their location, but in different ways. The most important result was that for some combinations of noise and signal positions, the response to the signal was slightly enhanced compared to the response to the signal alone. They tried to explain this phenomenon via masking level difference (MLD) but realised that these two events are fundamentally different from each other, as
the MLD is based on low frequency inputs and they used high frequencies. However, they emphasised that detection of a signal in the presence of the noise is better when the masker is spatially separated than when they are in a same location (Yost and Sfeft, 1993).

The IC is the last processing stage in the brainstem and receives binaural information. Most of its output is determined by inhibitory patterns. Gooler et al. (1996) highlighted that binaural inhibition may be responsible for the direction of motion of a source. It seems likely that these neurons in the IC play a role in spatial processing of the target signal intermingled with other sources of sounds. Changing the direction of sounds gives rise to different binaural cues. Gooler et al. (1996) found that by occluding the ipsilateral ear, and therefore decreasing the strength of binaural inhibition, direction-dependent frequency selectivity was eliminated (Gooler et al., 1996).

Lastly, the significance of the IC can be examined via transection of its brachium, the projections to the medial geniculate body (MGB) (Casseday and Neff, 1975). After unilateral abolition of the brachium, cats experienced a deficit in retention, relearning or accuracy of localisation. This suggests that normal behavioural performance in sound localisation depends on the intact information from both sides of the auditory system. With bilateral disconnection of the IC of three cats, two could relearn to localise, but with difficulty. The researchers suggested that some of the information needed to arrive at the level of cortex for localising behaviour might be relayed via pathways outside the primary auditory pathway.

To recapitulate, the IC behaves as an auditory hub in the auditory pathways, which converges processed information from the DCN, LSO, MSO, and LL to create outputs for the next processing stage. In a co-located condition, it seems more likely that the IC units converge the binaural information, which includes spatial cues in the horizontal plane, with the monaural information. The monaural information is directly received from the DCN (remember that the SNR of this information is already increased by inter-neural networks and other inhibitory neural units in the DCN) which conveys spatial cues in the vertical plane. This contrasting information associated with inhibitory neural networks in the ipsi- and contralateral ICs might improve the SNR. In a separated condition, we have two advantages. First, because for frontal sounds, the neural representations of the signal in one ear are synchronous with those from the other ear. Second, with separated noise to the sides, the neural population that is responsible for processing acoustic signals in 0 degree
azimuth is protected from the destructive effects of distractors and hence processes the signal more accurately. In the LL region, the units refine and might enhance the pre-encoded ILD and ITD cues, and convey them to the next stage of processing: the medial geniculate body (MGB).

2.4.2.5. Medial geniculate body (MGB)

The medial geniculate body (MGB) is the specific part of the thalamus that responds to auditory inputs received from afferents from the IC. The MGB dynamically shapes these representations and projects them to the auditory cortex. Due to its close reciprocal connections with the auditory cortex (AC) the MGB and AC are grouped together as a functional unit (Pickles, 2008).

The MGB has three divisions: the ventral, dorsal and medial. The ventral division is classified as a part of the lemniscal pathways because it is specialized to process auditory inputs only. They project to the primary auditory cortex (AI), the anterior auditory field (AAF) and the posterior auditory field (PAF). In contrast, the dorsal and medial parts are categorised as extralemniscal, because they integrate auditory and non-auditory information together and mediate the meaning of sounds in human beings and other species. The medial MGB receives inputs from somatosensory, visual, the spinal cord and auditory inputs from the ICC, the ICX, the SOC, the contralateral DCN and different parts of the VCN and projects to the core of the AC and the belt as well (Bazwinsky et al., 2003). The dorsal division of the MGB receives inputs from the ICD and the somatosensory system and projects to the core and the belt of the AC (Winer, 1985). It is worth noting that at the level of MGB and auditory cortex, the lemniscal and extralemniscal pathways are renamed “core” and “belt”, respectively.

The main projections from the IC to the ventral MGB originate from the ipsilateral ICC (Gonzalez-Hernandez et al., 1991). In some species with interneuron units (e.g., cat and primates), the IC fibres synapse both to interneurons and to those that are heading towards the auditory cortex (Guillery, 1995, Majorossy and Kiss, 1976). The ventral cells receive a considerable volume of inputs including inhibitory inputs from the interneurons, excitatory and inhibitory inputs from the IC, excitatory inputs from the AC, and inhibitory inputs from the reticular thalamic nucleus (Figure 8) (Winer et al., 1999). This arrangement of
innervation particularly from the IC neurons stimulates both thalamocortical and interneuron dendrites simultaneously and because of connections between the interneurons and thalamocortical neurons (Huang et al., 1999), the interneurons quickly suppress thalamocortical neurons’ activity via a feedforward inhibition (Sherman and Guillery, 1996). This real time network adjustment could regulate and shorten the duration of excitation in the MGB units resulting in a precisely timed control of fibre outputs, which is very important in suppressing echoes, reverberation and even would be useful in representing rapidly changing stimuli or pitch cues. This mechanism might explain to some extent how the auditory system benefits from better-ear glimpsing.

Figure 8. One example of a neural cell in the ventral part of the medial geniculate body (MGB). It receives inhibitory inputs from the IC, the AC and the interneurons of the nucleus (Winer et al., 1999).

Apart from the high resolution of frequency selectivity at the MGB, due to a multiplicity of inhibitory inputs, the lateral inhibitory interactions play a role in enhancing locational cues to elevation (Samson et al., 2000). These cues, as mentioned earlier in Section 2.4.2.1, are established by the spectral peaks and troughs of sounds impinging on the pinna, torso, head and neck and are primarily extracted by the inhibitory networks in the DCN (Imig et al., 2000) and first enhanced in the ICC (May et al., 2008), and then augmented in the MGB (Samson et al., 2000).

A dynamic relationship exists between the AC and the MGB for modifying required information. The complex inhibitory network for adjusting the acoustical inputs also aids the
MGB in both enhancing the signal and minimizing distracting noise when both occur at the same time. In some species, such as bats, there are neurons that are sensitive to the duration of auditory stimuli or to delay between stimuli. These characteristics may play a role in processing the signal and noise in different ways when they are spatially separated.

Although the MGB acts as a relay station for all the output of the IC, the MGB receives its inputs mostly monaurally from different nuclei in the brainstem. It may play a role in monaural spatial processing via refinement and enhancement of spectral cues. As a result, it might not play a role in improving binaural spatial processing. However, it seems likely that the two levels of binaural processing in the auditory pathways in the level of the brainstem including the SOC (MSO and LSO) and the IC (ICC) provide the binaural spatial cues needed by the auditory cortex. Thus, the MGB might play two roles: first, it may improve the SNR of the processed spatial information to allow for better understanding the target signal and second, it may process spectral cues more finely due to its neural inhibitory networks and convey this refined information to the auditory cortex for the final stage of processing.

2.4.2.6. Auditory cortex (AC)

The auditory cortex (AC) can be divided into three parts or regions. The precise anatomical locations and nomenclature of these divisions differ depending on the classic or modern definitions. As both definitions are still common, they will be both described below.

According to the modern definition, the AC comprises of three general regions: the core, the belt, and the parabelt. The core region is specialized for processing the basic characteristics of the acoustic signal. The belt and the parabelt are responsible for the detection of more complex features such as the determination of sound location. These multiple cortical regions are assumed to contribute to parallel processing, but also exert preference, analysing specific aspects of the auditory inputs. The auditory representation of the signal at the level of the AC is an auditory object (Pickles, 2008) formed on the basis of the signal’s acoustic cues, including its location.

In the classic subdivision, the auditory cortex (AC) is divided into three separate parts: the primary (AI), secondary (AII) and tertiary auditory cortex (AIII). These structures
are formed concentrically around one another, with the primary AC in the middle and the tertiary AC on the outside (Semple and Scott, 2003) (Figure 9).

The patterns of activity in the neural population of the core, belt, and parabelt are various. Cells in the posterior auditory field (PAF) of the core region commonly have inhibitory sidebands that flank the excitatory response area on both sides and a large proportion has multiple inhibitory bands. These neurons are thought to be involved in the temporal and spectral integration of complex signals. In addition, neurons in the PAF are more sensitive to sound location than neurons in the AI (Stecker et al., 2003). The belt
regions are involved in processing of the complex features of the signal, such as the
determination of sound location.

Borderline regions lie between the core and the belt where units behave in a
different way. Stecker et al. (2005a) found binaural neurons in an area between the core
and the belt (dorsal zone) of the AC in the cat, which have complex frequency tuning with
multi excitatory and inhibitory inputs for the sharp spatial selectivity of sound in the
horizontal plane with possible high frequency sensitivity. This finding suggests that these
cells might have a role in the spatial representation of a signal. Further areas reported in the
belt area of a cat were the partially overlapping visual, somatosensory, and auditory fields
that are involved in sound localization and in the processing of complex stimuli (Clarey and
Irvine, 1990). However some cells are exclusively specialized for the localization of sound
sources particularly the later part (Tian and Rauschecker, 2004).

In monkeys and in other mammals, the timing of individual neuron discharges at the
core area is periodic and depends on ITD. These neurons are able to precisely follow the
stimulus frequency below about 2 kHz (Rupert et al., 1963, Tasaki, 1954) and because they
represent ITD cues, known as binaural summative. In the core area, each neuron has a
characteristic delay, and encodes a particular ITD so that a set of them represents the entire
range of ITDs received by the listener. Within the core, neurons that represent high
frequency information are sensitive to ILD cues. In comparison to the ITDs, the function of
firing strength and timing of ILD-sensitive neurons is sigmoidal, meaning that, depending on
sound source location, the largest ILD occurs in the angles around the frontal hemisphere
(Figure 10) (Park et al., 2004). It is reported, however, that there are neurons in the AC of
the cats that are sharply tuned to binaural stimuli coming from the frontal angles (see
Section 2.4.2.2). These units are known as central units, and they have a best response
when the source is on the midline. Their responses decrease progressively to zero for posi-
tions more than 20˚ away from the midline. There is another batch of neurons which are
marked as omnidirectional units which discharge independently of the sound source
Chapter 2: Spatial processing and its disorder

Figure 10. The three schematic graphs illustrate idealized spike rate versus ILD functions for different neurons. When the level of the sound at the contralateral ear is greater than the level at the ipsilateral ear, the spike rates are high. The shaded area indicates the range of ILDs around 0 degrees azimuth where the greatest change in spike rate occurs (Derived from Musiek and Chermak (2007)).

There is considerable evidence based on lesion studies that the AC in humans (Cunningham et al., 2002, Zatorre and Penhune, 2001) and animals (Jenkins and Merzenich, 1984, Malhotra and Lomber, 2007) plays a role in localisation of sound and spatial processing, although the underlying mechanisms are not well described. Figure 11 illustrates the spatial tuning curve of a single unit in the AI area, representing auditory space in the azimuth plane. This kind of spatial receptive field gives rise to the panoramic nature of spatial processing which covers a wide range of the spatial acoustic field (Middlebrooks et al., 1994). The important role of the AC for representing sound sources in auditory space has been documented by several investigations that found deficits in sound source localisation following damage to the AC. Neff (1968) demonstrated that cats with large bilateral injuries were not able to determine sound sources. Further investigations showed that a lesion of one side had a negative influence on the localisation of sound sources on the hemifield of the opposite side in space. For example, Jenkins and Merzenich (1984) indicated that after unilateral damages of the AI in cats, they had profound deficits in localizing sound sources on the opposite side to the damage. However, this could occur even if all the spatial...
processing was carried out in the brainstem prior to the signal being transmitted to the damaged location in the auditory cortex.

Figure 11. Spike patterns of a single unit in the core auditory cortex. Each point represents the time of occurrence of a spike that is evoked by one stimulus presentation at the horizontal plane. Negative azimuth angles indicate the left side of animal which is contralateral to the place of recording (Middlebrooks et al., 1994).

There is a growing hypothesis that spatial processing in humans is implemented by the contribution of two right- and left-perceptual channels, each with tuning to the azimuth plane of one hemifield which overlaps in the midline (Boehnke and Phillips, 1999, Phillips and Hall, 2005). The neurophysiologic evidence has also been provided in an animal experiment (Stecker et al., 2005b). When the target speech stimulus is presented concurrently with noise in the same hemifield, there is a 1.3 dB spatial advantage if there is 90° separation between speech and noise. This perceptual advantage increases up to 8.6 dB, if speech and noise are located in two different hemifields, i.e., on the opposite side of midline (Phillips et al., 2003).

It means that a significant spatial advantage occurs when speech and signal are located in different channels such that they are available for selective scrutiny, but not when there is a smaller separation between them. Although this approach sheds some light on how the auditory cortex represents auditory space in the frontal hemisphere, it is unable to reveal how separation is achieved in complex environments, such as a cocktail party where there
are many spatially separate sources. This problem might be solved by turning the head towards the intended source, to take advantage of frontal priority once again. In this hypothesis, it may be predicted that there are two external factors affecting how the auditory cortex processes auditory space: frontal two-perceptual channels (the right and left hemifields) and the signal to noise advantage at the ear near the signal. In other words, the ability of people to localise and recognise the source of interest and maintain perceptual separation of two or more competing stimuli are based on dynamic strategies including perceptual channels and spatial location of stimuli.

This section provided a detailed overview of the individual components of the auditory system, ranging from the periphery to the auditory cortex, together with their relevance for spatial processing. As this thesis will employ electrophysiological methods to objectively evaluate spatial processing, the next section will describe several electrophysiological techniques that could be used for this purpose.

2.5. Auditory-evoked potentials (AEPs) and their potential for assessing spatial processing

This section introduces the concept of using electrophysiological recordings to detect the presence/function of spatial processing, or spatial release from masking (SRM). This involves recording of auditory evoked potentials (AEPs) from different levels within the central auditory nervous system (CANS). It will describe which types of AEPs could potentially be used to evaluate spatial processing by looking at their use in similar research investigating the effects of ITDs, ILDs and binaural processing on AEPs.

AEPs are electrical activities of the neural populations in the auditory system to auditory stimuli such as clicks, tone bursts, complex tones, or speech. These activities can be detected through electrodes placed at specific locations on the scalp (Ruth and Lambert, 1991), and filtering of the responses to identify electrical activity that is time-locked to the stimulus. Most AEPs have extremely low amplitudes (in the order of microvolts, \(\mu V\)) hence cannot be seen directly in the routine electroencephalography (EEG) recording. Separating these tiny waveforms from the relatively high amplitude background brain activity (e.g.,
muscle contractions, eye artefacts and other internal and external noise sources) requires specific response analysis procedures (Harinder et al., 2010).

Two basic analysis procedures are necessary for detecting AEP waveforms: signal amplification and signal averaging. The amplification of the recorded signal magnifies the signal (tens of) thousands of times to prepare it for averaging (Chiappa, 1997). For instance, a gain of 100,000 to 150,000 is applied to ABR signals. Signal averaging relies on the responses being time-locked to the stimuli, i.e. they always come at the same interval time after the stimuli. This allows the individual AEP responses to be added together and then divided by the number of responses. The background noise, however, is not time-locked to the stimuli, which causes it to be decreased due to destructive interference. This increases the signal-to-noise ratio with which the response is measured.

Any successful recording of AEPs depends on appropriate recording protocols (Figure 12). These factors comprise participant parameters (age, gender, body temperature, state of arousal, drug effects, and artefact sources), stimulus parameters (intensity, rate, polarity, and duration), and acquisition parameters (type of electrodes used, electrode placement, time epoch analysis, filtering settings) (Guerreiro and Ehrenberg, 1982, Jewett and Williston, 1971).

AEPs can be classified based on their latencies (Figure 13). Latency is defined as the time of occurrence of the AEP wave component after a stimulus, or the time interval between two AEP wave components, and is expressed in milliseconds (ms) (Hall, 2007).
In this thesis, it is particularly interested in the use of AEPs that occur above the level of the cochlea. Given this, a discussion of the auditory brainstem response (ABR), the frequency following response (FFR), the middle latency response (MLR), and the cortical auditory evoked potential (CAEP) and the AEPs evoked beyond the auditory cortex will be provided. In the following section, each AEP will be reviewed briefly and the anatomic origin of each response will be discussed. An overview will be provided of spatial processing and some theoretical points regarding the characteristics of AEPs that might be used to record possible representation of SRM. To conclude, the rationale behind the selection of the AEPs, which have been employed in this thesis, will be explained.

### 2.5.1. Electrocochleography (ECoChG)

We acknowledge that electrocochleography (ECoChG) (Figure 14) is an electrophysiology test that measures signals arising from the cochlea, and the auditory nerve and represents initial processing of an acoustic stimulus monaurally; hence, it cannot directly provide information about binaural spatial processing of a target signal. However, it might provide some information regarding cochlear frequency analysis that affects the spatial processing of a target signal in higher centres.

ECoChG occurs within the first 2-3 ms after an onset of a short stimulus. It includes three components, cochlear microphonic (CM), summation potential (SP), and action potential.
Chapter 2: Spatial processing and its disorder

potential (AP). The CM is the first component, which is an alternating current potential that imitates the waveform of the stimulus and arises from hair cells (Dallos, 1973, Sellick and Russell, 1980). The SP arises from the nonlinear characteristic of transduction. The AP is the compound action potential caused by synchronous firing of a number of auditory nerves (Dallos, 1973, Eldredge, 1974). To the author’s knowledge, no study has considered the impact of spatial processing on ECoChG response empirically, which theoretically could occur because of efferent control of the cochlea. At least some physical interactions arising from energetic masking of a target by distractors could be identified via ECoChG.

![Graph showing an electrocochleogram waveform](image)

**Figure 14. Example of an electrocochleogram waveform (Hall, 2007).**

### 2.5.2. The auditory brainstem response (ABR)

The auditory brainstem response (ABR) is an electrical activity from the brainstem evoked by a brief stimulus. It is made up of several waves (I, II, III, IV, V, VI, and VII) (Guerreiro and Ehrenberg, 1982). This potential can be described by two important indices: the latency and the amplitude of each peak. These indices are affected by the speed of transmission and the number of neurons simultaneously activated, respectively. This response represents, as shown in Figure 15, the neuroelectrical synchronous activity of the eighth cranial nerve fibres (the distal portion and proximal), the cochlear nucleus, the trapezoid body, the superior olivary complex, the nucleus of lateral lemniscuses, the lateral
lemniscuses tract, the inferior colliculus, and the thalamus (Chermak et al., 1999, Chiappa et al., 1979, Guerreiro and Ehrenberg, 1982, Sininger and Hyde, 2009).

2.5.2.1. Possible generators underpinning the ABR

The locations of the neural generators of the different components of the ABR are still partly a matter of debate. This debate is partly due to the contribution of multiple regions in the brainstem to individual ABR components. The generators of early components, wave I and II, however, have been labelled with more confidence than the later components, waves III-VII. Wave I certainly originates from the VIII cranial nerve (Jewett and Williston, 1971, Sohmer and Feinmesser, 1967). This wave corresponds to the compound action potential (AC) recorded by the ECochG and is generated in the peripheral (distal) part of the VIII cranial nerve (Luders et al., 1988, Grandori, 1986, Ino and Mizoi, 1980). The proximal, intracranial, part of the VIII cranial nerve, on the other hand, contributes to wave II ABR. Apart from the intracranial part of the VIII cranial nerve which has been shown to be involved by direct recording from this nerve during surgery, the ipsilateral rostral part of the medulla is also suggested to contribute to wave II (Møller et al., 1988, Møller and Jannetta, 1982, Møller et al., 1995). At the timing of wave II, the trapezoid
body (TB) likewise has shown significant activity in animal studies (Wada and Starr, 1983). The generator of wave III is less certain but it is believed that this wave originates from the auditory pathways within the caudal pons both ipsi- and contralaterally (Achor and Starr, 1980, Legatt et al., 1986, Grandori, 1986, Scherg and Von Cramon, 1985). This finding has been verified by a magnetic resonance imaging (MRI) study (Markand et al., 1989). In parts of this study, it was reported that lesions at the level of midbrain did not affect waves I-III. However, lesions within the pontine region are primarily associated with prolongation of I-III interpeak latency. An interpeak latency is defined as the differential in time between two successive peaks. The location of wave IV generators remains mainly uncertain. Two neural generators are suggested in the literature, the superior olivary complex (SOC) (Møller et al., 1995, Moore, 1987) and the inferior colliculus (IC) (Lev and Sohmer, 1972, Scherg and Von Cramon, 1985). Regarding wave V, it has been suggested that the generator of this wave is possibly located in the most rostral part of the lateral lemniscus (LL) or in the IC or both (Hashimoto et al., 1981, Møller et al., 1995). Clinical trial studies in patients with unilateral midbrain injuries have shown an attenuated or absent wave V (Markand et al., 1989). The generators of waves VI and VII are unclear due to inconsistency of the waves. Thus, different neural populations have been suggested in the literature as the origin of these AEPs components, including the medial geniculate body (MGB) and the midbrain (Curio and Oppel, 1988, Hashimoto et al., 1981, Møller et al., 1995).

2.5.2.2. Using the ABR to study spatial processing at the level of the brainstem

ABR recordings have been used to investigate the functioning of the auditory system at the level of the brainstem in processing of acoustic signals distributed spatially using different cues. These cues comprise interaural time differences (ITD), interaural level difference (ILD), and monaural spectral cues. To investigate the effects of ITD and ILD cues on ABR components, these cues have been applied to acoustic signals either individually or in combination with each other. Furthermore, to simulate an everyday listening situation and to assess the ability of the auditory system to process such situations, a three-dimensional acoustic environment can be developed by incorporating all spatial cues and presenting these under headphones.
Different studies have investigated ITD and ILD cues to identify how, when and where these cues are reflected by changes in ABR components (Brantberg et al., 1999, Furst et al., 1985, Gerull and Mrowinski, 1984, Jones and Van der Poel, 1990, McPherson and Starr, 1995, Riedel and Kollmeier, 2002). These studies often have used binaural interaction component (BIC), or binaural difference (BD) potentials, calculated by taking the difference between binaural and summed monaural AEPs. This particular component, labelled as DN1, is the most consistent ABR BD, and typically occurs at a latency of ABR wave V or just after (Ito et al., 1988). These studies demonstrated that increasing the ITD was associated with an increase of DN1 latency. In the case of DN1 amplitude, however, there is some controversy in the literature. While some studies show unchanged amplitudes for decreasing ITDs down to around 1 µs (Brantberg et al., 1999, Furst et al., 1985, Jones and Van der Poel, 1990), others reveal close correlations between DN1 amplitude and ITD (McPherson and Starr, 1995). When it comes to ILD, investigators identified that with increasing difference in ILD, amplitudes and latencies of the ABR decreased and increased respectively (Furst et al., 1985, McPherson and Starr, 1995).

Pratt et al. (1997) tried to identify whether ILD and ITD cues were analysed independently in the brainstem or whether these spatial cues were analysed in combination (based on the disparity in time between the two ears). They used ABR to monaural and binaural clicks with different ITDs and ILDs. They found that these cues were analysed independently in the brainstem at least at the levels of the generators that waves II and III arise from.

Furst et al. (1985) investigated sound lateralisation utilising ILDs and ITDs in the light of binaural fusion, which were reflected in a binaural component through ABR recording to click stimuli. In this research, they used β nomenclature for labelling the first major peak of the binaural component. They reported that as the ITD increased, to some extent, the β amplitude was relatively constant when the binaural stimulus was perceived as a clear fused image. However, the β amplitude dropped immediately when the fused image stopped shifting to the side, losing its quality and becoming blurred in the process. In the case of increasing ILD, as the fused image shifted towards the ear receiving the louder stimulus, the β amplitude started to decrease steadily. When the ILD reached a certain value, the β amplitude dropped to zero. Based on these results, it seems likely that introducing target sound stimuli at ITD and ILD 0° azimuth creates a clear fused image, which can be, reflected
via ABR waveform components. However, when these stimuli were presented together with some distractors concurrently, two fused images will be perceived, one for the target sound and the other for the distractors. The target image loses its quality and becomes blurred due to mixing or overlapping with the distractors’ images. As a result, respective ABR waveforms are negatively influenced. By shifting the distractors to the sides however, the quality of the distractors’ image vanishes and the quality of the target image becomes clear. This results in the improvement of the ABR waveforms.

Using virtual acoustic environments created by combining all spatial and spectral cues to investigate the effect of spatial cues on the generation of the ABR (while considering both binaural potentials and BD), Junius et al. (2007) reported that with increasing laterality of the sound source, ABR amplitudes and latencies decreased and increased respectively. They concluded that binaural interaction at the level of the brainstem is commonly influenced by spatial cues.

### 2.5.2.3. The ABR might correlate with spatial processing

In this thesis, ABR wave V will be evaluated as an objective measure of spatial processing because it is a common component in clinical applications as it is relatively easy to record. The generator of this wave is assumed to be located in the termination of the LL as it enters the IC, and the IC itself (Buchwald and Huang, 1975, Hashimoto et al., 1981, Møller et al., 1995).

Due to the contrasts between the neural information coming from the ipsi- and contralateral auditory pathways at the IC level, the processing of all spatial cues and their combination in the IC, and its contribution in generating the ABR wave V, it is reasonable to investigate whether recording of this wave could be used to obtain information about the neural structures underlying SRM. Looking more specifically to our protocol which will be introduced in Chapter 4, it is hypothesised that in the separated condition, where the distractor signals are moved to the sides, the amplitude and latency of ABR wave V increase and decrease respectively. If this hypothesis is correct, this might suggest that the ABR demonstrates SRM.
2.5.3. The frequency following response (FFR)

The frequency following response (FFR) is a steady state response and demonstrates sustained neural activity, which is integrated over a population of neural elements. The FFR can be recorded during practically the entire length of a stimulus presentation (Figure 16). It is phase-locked to the individual cycles of a stimulus waveform or envelope of periodic stimuli (while ABR is a transient and onset response). It is generated by presentation of low frequency stimuli and exhibits the period of the stimulus fundamental frequency (F₀). More importantly, FFR represents information about the spectrum and pitch of speech sounds as it follows the periodicity of stimulus (Krishnan, 2007). Figure 16 demonstrates the grand average of a FFR to 1000 /u/ stimuli (fundamental frequency F₀= 110 Hz) with duration of 250 ms presented at 60 dB SPL in time (left panel) and frequency (right panel) domains. The most robust amplitude in the frequency domain is the FFR F₀ at 110 Hz followed by its higher harmonics.

![Figure 16. An example of a FFR waveform in time and frequency domains (from the current study).](image)

2.5.3.1. Possible generators underpinning the FFR

The FFR was first described by Worden and Marsh (1968) in animal studies. Subsequently, many research studies have attempted to identify the neural generators underlying this response. Most researchers have concluded that the neural generators for the FFR are within the brainstem (Marsh et al., 1974, Marsh et al., 1975, Marsh and Worden, 1968, Moushegian et al., 1973, Smith et al., 1975). It seems possible that the FFR
can be recorded from all nuclei in the brainstem including the CN, the SOC, the TB, the LL, and the IC (Marsh et al., 1972). However, it is the type of stimulus and intensity level that determines which neural populations are phase-locked to the stimulus and reflect as the FFR. Using higher intensities results in changing the morphology of the FFR because they produce shorter latencies and hence new early EP components. Identified that, using an electrode montage between ipsilateral earlobe and vertex and high intensity monaural tone bursts, the recorded responses include a complex of short-latency cochlear microphonic (CM) components superimposed with a longer-latency neural component. At moderate intensities (60-70 dB SL) from intracranial recording in animals, Smith et al. (1975) reported some evidence that the IC seems to be the primary source of scalp recorded FFR. More evidence has been introduced by Sohmer et al. (1977) who recorded EPs in a patient with lesions in the region of the IC and demonstrated that no measurable FFR could be detected. Thus, all the evidences pinpoint the IC as the neural generator primarily causing the FFR.

Two FFR components have been identified, FFR\(_1\) and FFR\(_2\). These components are differentiated from each other by simultaneous recording of the FFR to low frequency pure tones in vertical and horizontal electrode arrays (Stillman et al., 1978). FFR\(_1\) was equal in both electrode arrays and had larger amplitude at higher levels. FFR\(_2\) was more pronounced in the vertical array and at lower intensity levels. It has been suggested that the horizontal array originates from the VIII cranial nerve while the vertical array originates from the central brainstem (Galbraith et al., 2000, Galbraith, 1994, Møller et al., 1988, Smith et al., 1975, Stillman et al., 1978).

Stimulus frequency determines the region of the auditory pathways responsible for generating the FFR. The upper frequency range for phase-locking decreases as the FFR originating location moves toward higher centres within the CNS (Langner, 1992). The frequency range for the CN is about 5 kHz, for the MSO about 3-4 kHz, and for the LL and IC about 1.5-2 kHz (Starr and Hellerstein, 1971). Furthermore, with increasing stimulus frequency and the number of harmonics in a complex tone, the FFR amplitude decreases accordingly for synthetic speech sounds, two-tone stimuli and tone bursts (Gardi et al., 1979a, Greenberg et al., 1987, Krishnan, 1999, Krishnan, 2002, Pandya and Krishnan, 2004).
2.5.3.2. Using the FFR to study spatial processing at the level of the brainstem

Two principal approaches to studying the capacity of the auditory system in separating a target source from other sources in complex acoustic environments have been described in the literature. These two approaches are: (1) binaural unmasking, which is assessed with the binaural masking level difference (BMLD) and; and (2) spatial unmasking which is enabled by more realistic interaural differences in spatial cues, ILDs and ITDs along with other spatial cues associated with spatially separated signals and maskers.

The first approach creates a BMLD by introducing a phase reversal of targets or distractors interaurally. This phase disparity results in the improvement of detection of the target signal, the phenomenon which was first described by Licklider (1948). Hirsh (1948a) and (1948b) reported that if the phase of the target signal in the presence of background distractors in one ear is inverted (π radians), the reception threshold of the target signal decreased (improved) by 12-15 dB. This decreasing of the target threshold is known in the literature as the BMLD. It is suggested that BMLD might reflect binaural processing characteristics in the MSO at the brainstem level (Jenkins and Masterton, 1982).

The BMLD has been investigated by Wilson and Krishnan (2005) utilising FFRs in human participants. It has been demonstrated that the FFR amplitudes were significantly increased in both $S_0N\pi$, (homophasic signal and antiphase noise) and $S\piN_0$ (antiphase signal and homophasic noise) in comparison with $S_0N_0$ (homophasic signal and noise) condition. These results imply that changing the phase of the signal or noise releases the signal from the masking effects of noise. Binaural unmasking in the rat amygdala was investigated by Du et al. (2009a) utilizing 150 ms chatter and a 150 ms tone complex in the presence of a broadband noise via FFR. The results showed that introducing an ITD between the chatter and maskers, while the ipsilateral target chatter had advantage in time compared to contralateral target chatter in the presence of interaurally correlated masking noise, resulted in enhancing the FFR amplitude and hence unmasking the FFR. In another study, the same investigators reported that the amplitudes of the FFR in the IC neural population to the target chatter in a rat were enhanced by inserting an ITD between the target chatter and the noise masker (Du et al., 2009b).
The second approach involves the introduction of ITDs and/or ILDs, either under headphones or in the free field environments using loudspeakers. Localisation refers to the perceived location of images outside of the head, the location of which can be varied by altering the ITDs and ILDs of the introduced sounds. In lateralisation experiments, in contrast, sound images fuse inside the head and can be moved intracranially by changing the ITDs and ILDs of the sounds. Clark et al. (1997) reported that lateralising tonebursts of 450 Hz with ITDs of 0 and 660 µs evoked larger FFR amplitudes for the 0 µs difference than for the 660 µs difference. To further evaluate different ITDs using lateralisation, the effects of ILD on the binaural interaction component (BIC, subtracting binaural evoked potentials from summed monaural evoked potentials) in the FFR (FFR-BIC) has been evaluated using a 500 Hz toneburst at different intensities. The FFR-BIC decreased with systematic increase of the ILD (Krishnan and McDaniel, 1998). Ballachanda and Moushegian (2000) using ILDs at different intensities and ITDs at different timings simultaneously showed that when one ear is ahead in time and louder in intensity (i.e., in concert) and ahead in time but quieter in intensity (i.e., in opposition), significantly different FFRs are created. FFR amplitude decreases when both cues are in favour of one ear (i.e., in concert). This amplitude was relatively constant when each ear gained only from one cue (being either ahead in time or louder in intensity).

2.5.3.3. The FFR might correlate with spatial processing

As already discussed there is a general agreement that FFR recordings from the human scalp represent neural activities at the level of brainstem (Gardi et al., 1979a, Stillman et al., 1978, Smith et al., 1975, Sohmer and Pratt, 1977), particularly from the IC neural population. Because it receives spatial information from the lower brainstem, the IC encompasses the neural generators that are able to process and encode spatial cues, which are important for enhancing the target and suppressing the distractors (i.e., SRM). Activating these neural generators in a way that they can reflect SRM depends closely on the type of stimuli that are used to evoke their responses.

While the mechanisms underlying localisation and lateralisation of a single-sound source using the FFR have received appropriate consideration in the literature, the question how the auditory system extract the target signal among multiple sound sources, or in
general in everyday acoustic environments, has not been addressed yet. To address these questions, it is necessary to study how the auditory system separates a target source in the presence of distractors, and whether the FFR can reflect the neural mechanisms undertaking this processing. To meet this aim, it is a prerequisite to simulate a three-dimensional (3-D) acoustic environment in which all basic and critical spatial cues including ILD, ITD, and spectral cues are considered.

In this thesis, a vowel sound and a tone complex with low fundamental frequencies and moderate intensities will be employed to correspond with the findings in the literature.

In this thesis, ecologically relevant acoustic environments are simulated using head related transfer function (HRTFs). Two spatial conditions are considered, spatially co-located and spatially separated. In the co-located condition, where both target and distractors are delivered at 0° azimuth, it is assumed that the FFR fundamental frequency (F0) amplitude is negatively influenced by masking effects of the distractors. This is similar to BMLD in the S0N0 condition and to the study of FFR binaural unmasking conducted by Du et al. (2009b) in the masking condition. In the separated condition, on the other hand, the target signal is presented still at 0° but the distractors are presented 90° off the frontal location.

It is hypothesised that the FFR F0 amplitude will be larger in the separated condition and hence reflects SRM. This condition is somewhat similar to Du et al. (2009b) experiment who introduced a disparity between the inter-aural phase of the target sounds (mouse chatter) and that of the masker. However, although the paradigms that will be utilised in this thesis are partly similar to these studies, there are some fundamental differences between them. In the paradigms used in this thesis, all critical spatial cues including ITD, ILD, and spectral cues will be included when evaluating spatial separation of the distractors from the target. In both cited studies only ITD cues were used, which limited them to lateralisation of sound. Thus, regardless of some similarity, it is hypothesised that the neural mechanisms underpinning the generation of the FFR that will be recorded in our experiments will likely be different from those reported in the cited studies.
2.5.4. The middle latency response (MLR)

Because of more uncertainty regarding the generators of the middle latency response (MLR), using of MLR in the current experiments investigating SRM has not be made. However, it is briefly reviewed for completeness.

The middle latency response (MLR) is an evoked response that occurs following the ABR from 12-50 ms after stimulus onset (Figure 17). The components of this response are Na, Pa, Nb, and Pb. The neural generators of these components are still debated in the literature. However, it is evidenced that Pa at least partly arises from the superior temporal gyrus in the auditory cortex (Goff et al., 1977, Lee et al., 1984). The Na component mainly originates from the midbrain with more contribution of the IC. For Pb, on the other hand, it is hypothesised that several anatomic regions contribute to generation of this component such as planum temporal, and hippocampus (Woods et al., 1987).

Sound localisation in both azimuth and elevation planes have been investigated by Polyakov and Pratt (2003) utilising MLR. In this study, they used click stimuli in nine virtual spatial locations responses and analysed the binaural interaction components (BICs). BIC is calculated by subtracting binaural response from the summed monaural responses. They found a significant decrease in magnitude of Pa2 in backward positions and of Nb in the back and the front positions in the midsagittal plane compared with directly above the head (maximum elevation). In the coronal plane, Pa magnitude decreased in response to right-horizontal location (no elevation) compared with directly above the head (maximum elevation). They also found significant effect of spatial location in midsagittal plane in backward position compared with other locations on latency of Nb component. They attributed these changes in the magnitudes and latencies of the MLR BICs across different spatial locations to spectral tuning in binaural-sensitive neurons in primary auditory cortex, underpinning processing of spatial cues for sound localisation. In another study using magnetoencephalography (MEG), middle-latency and late potentials were recorded to determine whether ITD affects these responses (McEvoy et al., 1994). The results showed that neither the P30m nor the P50m was influenced significantly by ITDs. The N100 was enhanced significantly, when ITDs were in the favour of the contralateral ear compared to when the ITD was zero or in favour of the ipsilateral ear. It is also reported the source location for P50 was different from that of P30m and P100m. These results suggested that
the recorded components represented different aspect of auditory processing that occur while processing spatialised stimuli.

Taken together, the MLR reflects some aspects of spatial processing. To create a comprehensive aspect of “where” in the auditory system occurs, the MLR should be considered in future studies.

Figure 17. Example of a MLR waveform (Hall, 2007).

2.5.5. The cortical auditory evoked potential (CAEP)

Cortical auditory evoked potentials (CAEPs) are potentials that can be recorded from the auditory cortex in response to onset, offset and changes of a stimulus (Lightfoot, 2010). The different components of CAEPs can be identified based on latency. They are labelled as P1, N1, P2, and N2 and generally occur within 50 to 200 ms after presentation of a stimulus in adults (Figure 18). In children or infants these latencies are longer, where latencies of the P1 can vary between 50 and 300 ms depending on the age of the child (Purdy et al., 2006). The CAEP is strongly affected by internal factors including reaction to the stimulus context, attention to the stimulus or task required, and the state of the nervous system while stimuli are presented. Any changes of these parameters may cause components to (dis)appear in a waveform (Hall, 2007).

Apart from the internal factors that influence the components of the CAEP, stimulus parameters are also important. These parameters are type of stimulus, duration, intensity,
rate, and the number of stimulus repetitions. Moreover, the cortical waveform components are affected by acquisition parameters including amplification, analysis time, prestimulus time, sampling, number of sweeps, filtering, and electrode configuration. The amplitude and latency of the N1 and P2 components decrease and increase, respectively, with stimulus frequency. The latency of N1 component becomes longer as the onset time increases with greater increases at lower intensities (Onishi and Davis, 1968). However, if the onset time is between several milliseconds and around 50 ms, the amplitude of response is relatively constant and then decreases for longer onset times (Onishi and Davis, 1968, Skinner and Jones, 1968). The amplitude also decreases and the latency becomes longer if the intensity level of the stimulus decreases (Picton et al., 1977). The rate of increase of the amplitude-intensity function of CAEP components is not linear and tends to decrease with increasing intensity (Ross et al., 1999). Regardless of the acquisition and stimuli parameters, there are significant factors related to subjects that need to be considered in recording CAEP. These include age, gender, state of arousal and sleep, attention, auditory training and different types of drugs.

\[ \text{Figure 18. Example of a CAEP waveform (Hall, 2007).} \]

### 2.5.5.1. Possible generators underpinning the CAEP

In comparison with early and middle latency responses, the neuroanatomic regions and respective neural populations underlying the components of CAEPs are vast and various hence the contribution of these regions to the cortical components is a big matter of debate. The CAEP is thought to be generated by multiple regions in the auditory cortex (Stapells, 2002). These regions include the primary auditory cortex, auditory association...
areas, the frontal cortex, and subcortical regions. Davis et al. (1939) assumed that the generators of the cortical response were in thalamocortical areas. While verifying this statement, Picton et al. (1974) found that the generators of this response might also be in the association cortex of the frontal lobe. In a study to identify the generators of CAEP components, Vaughan and Ritter (1970) recorded CAEPs in six normal participants. They reported a clear change in polarity of cortical components at the Sylvian fissure region for latencies of 200 ms and below (Figure 19).

Figure 19. The reversal trend in polarity of CAEP components over the regions of the generator sites of the scalp (Vaughan and Ritter, 1970).

### 2.5.5.2. Using the CAEP to study spatial processing at the level of the auditory cortex

#### 2.5.5.2.1. Research focusing on ILD and ITD as individual cues

The body of AEP research into spatial lateralisation and localisation of sounds has usually focused on ILDs and ITDs as individual cues, without considering a natural acoustic environment (Butler, 1972, Gerull and Mrowinski, 1984, Polyakov and Pratt, 2003, Schröger, 1995, Schröger, 1996, Wambacq et al., 2007, Woods and Clayworth, 1985, Picton et al., 1991). In an experiment to examine the effect of localisation of sound sources on the auditory evoked response, Butler (1972) found a significant growth in amplitude (N1P2) when sounds alternated between two loudspeakers (with azimuths at either 270° and 315°,
270° and 0°, or 270° and 90°, or with elevations at 45° above and 45° below ear level). He attributed these results to a place theory of cortical organisation for auditory space perception. A considerable amount of AEP activation was found by Smith et al. (2010) and Wambacq et al. (2007) in both children and adults in a condition with an ILD of 20 dB in comparison with an ILD-absent condition. This experiment demonstrated a negative wave at 100 ms in adults, a positive wave at 200 ms in both, and finally a positive prolonged wave at 250 ms in children. Moreover, Picton et al. (1991), in an attempt to determine the effect of lateralisation of a sound on a human evoked potential, found an increase in amplitude with increasing ITD.

2.5.5.2.2. Separate processing of ITD, ILD and F₀?

Trying to identify whether ITD and ILD cues are processed by separated or integrated neural codes in the human auditory cortex, Edmonds and Krumbholz (2014) measured CAEPs in response to changes in sound lateralisation created by different ILDs and ITDs. Disparities in ITD and ILD resulted in different cortical morphologies. It was concluded that the auditory cortex has dual processing approaches toward laterality of a sound. While the auditory cortex integrates the codes for sound laterality, Edmonds and Krumbholz concluded that it also maintains independent information about ITD and ILD cues for the next stage of spatial processing.

It has been questioned in the literature whether changes in spatial location and pitch of stimuli are processed separately or together in the auditory cortex. Using tonal sequences with changing pitch and location, Warren and Griffiths (2003) showed that pitch-related activations induced more lateral and anterior activation of the planum temporal (PT) than the location-related activations. In another study, Barrett and Hall (2006) demonstrated that changes in spatial location largely induced activity in different regions in the auditory cortex (non-primary auditory cortex in the PT) than changes in pitch (Heschl’s gyrus and more anterior regions). However, some studies revealed a similar impact on the auditory cortex of changes in content of sounds and their location, suggesting some degree of overlap between processing of changes in spatial and non-spatial parameters. Smith et al. (2010) used fMRI to show the combined effects of spatial location and increasing the number of talkers in normal participants. The results revealed that increasing the number of talkers
modulated activities within the spatial PT regions. They concluded that sensitivity to spatial cues in the PT in the previous studies represented auditory source separation based on spatial cues rather than spatial processing per se.

2.5.5.2.3. Effects of types of masking (energetic versus informational) and attention

The effect on spatial separation of speech sounds in the presence of both energetic and informational masking using ITD cues and two attention and non-attention conditions has been investigated via CAEPs in normal hearers (Zhang et al., 2014). The results demonstrated that the effect of two-talker speech maskers on speech stimuli, represented by a cortical N1/P2 complex, was larger than other masking stimuli including steady-state-noise and amplitude-modulated speech-spectrum-noise. Further, shifting from non-attention to attention and co-located to separated conditions resulted in an enhancement of the N1/P1 complex only in the presence of speech maskers. Thus, the effect of attention and spatial separation on the early cortical representation of target stimuli is influenced by the texture of masking being used.

2.5.5.2.4. Approaches other than EEG to study spatial processing at the level of the cortex

The effects of spatial and spectral change on speech elements at the level of cortex has been studied in human normal participants via MEG (Du et al., 2011). This study showed that when the target and masker differed in spatial location and pitch, the negative peak around 140 ms after speech stimulus onset was enhanced compared to when they differed in only one cue or when they had similar spatial and spectral cues. In addition, the effect on the cortical responses when both types of cue were present was the sum of the effects that occurred for each type of cue in isolation. The authors concluded that the acoustic information combines linearly to improve perceptual separation between two simultaneously occurring sounds.

Processing of ILDs and ITDs has been investigated at the level of cortex through MEG, with good temporal resolution (Fujiki et al., 2002, Johnson and Hautus, 2010, Lewald et al., 2002, McEvoy et al., 1993, McEvoy et al., 1994, Tiitinen et al., 2006), functional
magnetic resonance imaging (fMRI) and position emission tomography (PET), with good spatial resolution (Bernstein, 2001). Furthermore, ILDs and ITDs have also been examined using MEG while employing HRTFs (Palomaki et al., 2000, Palomaki et al., 2002, Palomaki et al., 2005). They showed that both hemispheres are involved in processing of acoustic stimuli coming from different spatial locations. Nevertheless, the right hemispheric activities were stronger in comparison with the left hemisphere, which the authors interpreted as the right hemisphere being specialised in the representation of spatial cues. The peak amplitudes of responses were larger for sound sources coming from the contralateral hemifield than for sources from the ipsilateral hemifield. The effects of ITDs and ILDs therefore occur in relatively late stages of cortical processing.

2.5.5.3. The CAEP might correlate with spatial processing

In Chapter 4, the potential of CAEP components P1, N1, and P2 to objectively reflect spatial processing will be investigated. Due to the uncertainty of the exact generators of these components, non-specific generators of thalamocortical and cortical regions are assumed (Hall, 2007). In Section 2.4.2.5 and 2.4.2.6, the anatomy, physiology and the possible role of the MGB and the AC in spatial processing were discussed. Although the MGB is thought to have a role in processing spectral monaural cues, when it comes to the cortex all pre-processed information converges in this centre for the determination of the sound sources in an auditory space. It is hypothesised that the degree of SRM may be greater within cortical structures than in the lower structures if additional spatial processing occurs within the cortex. Therefore, in the second experiment of this thesis, it will be assumed that in separated conditions, the amplitude and latency of CAEP components might be increased and decreased respectively and hence indicate the effect of SRM.

In conclusion, the large body of electrophysiology research into lateralisation and localisation sheds light on the underlying neural mechanisms, and shows that any change in ILD and ITD cues could be traced temporally and spatially through these electrophysiological tools. However, the auditory mechanism behind spatial processing of the signal and the background noise under three-dimensional and naturalistic conditions (where ITD, ILD and spectral cues co-vary in an ecologically valid manner) still needs to be established.
2.5.5.4. AEPs originating beyond the auditory cortex

Beyond the time of occurrence CAEPs, there are several AEPs, which usually are categorised as event-related potentials (ERPs). ERPs that appear after CAEPs are P300 (P3), N400, and P600. Because these components usually are induced in combination together, conclusions about individual components are dependent on using pairs of paradigms differing in some way, such as using word vs. tone or performing different tasks such as attending or ignoring a stimulus.

The P3 response, also referred to as P300 as it occurs 250 to 500 ms after stimulus onset, is a cognitive response that is evoked by attending to a stimulus (Figure 20) (Polich, 2007). Two types of stimuli are needed to record P3, “rare” (target) stimuli, which are presented in 10-20% of the stimuli and “frequent” stimuli, which are presented most of the time (Sutton et al., 1965, Picton et al., 2000). P3 has a two components, P3a which is induced independent of attention to the rare stimuli and P3b, which is the conventional P3 response that is evoked only with attention to the rare stimuli (Squires et al., 1975). The origins of P3 are diverse and many anatomical regions in the brain (e.g., frontal lobe, auditory regions, and subcortical structure including hippocampus, and thalamus) contribute to this cognitive component.

The N400 is a negative deflection that occurs about 400 ms after a stimulus onset and is evoked by an unexpected (typically final) word in read sentences (Kutas and Hillyard,
1980), in response to a semantic violation. The generators of N400 are suggested in a wide-
spread collection of cortical areas including superior/middle temporal gyrus, the temporo-
parietal junction, and the medial temporal lobe, as well as some frontal regions (Halgren et al., 2002, Tse et al., 2007).

The P600 is language-based ERP and is evoked in response to grammatical or
syntactical errors. It is a positive deflection that appears at about 500 ms and reaches its
peak at about 600 ms, and continues for several hundred milliseconds afterwards (Hagoort et al., 1999, Gouvea et al., 2010). It is suggested that the origin of this cognitive component
is in the posterior temporal lobe, behind Wernicke’s area (Helenius et al., 2007).

Schröger (1995) conducted a study via ERPs (including mismatch negativity (MMN),
CAEP, and P300 responses) varying frequency and location both separately and
concurrently, in both an attended and non-attended condition. In the non-attended
condition, the results demonstrated that the MMN for the two-dimensional deviant (one
frequency and one location deviant) was equal to the sum of the two one-dimensional
deviant waveforms. This is likely the result of activity in distinct and independent neural
populations, one responsible for detecting changes in location and one responsible for
detecting changes in frequency. However, in the attended condition, which evoked N2b and
P300 ERPs, the response to the two-dimensional condition was less than the sum of the
responses to the one-dimensional conditions. This likely indicates that there are additional
neurons involved that respond when any acoustic component of the input changes. In
another study using MMN (Schröger, 1996), based on ILDs and ITDs, the representation of
these binaural location cues were somewhat processed in parallel but there was a distinct
representation of these cues at a cortical level.

The impact of spatial cues individually or in combination together using P300, N400,
and P600 has not yet been investigated. However, Romei et al. (2011) examined the effect
of multi-talker babble on cortical N400 ERP while processing word triples in quiet and in the
presence of background babble. They found no effect on N400 in the interior region but
enhanced neural processing negativity in anterior and central scalp regions in response to
the second word but not for the first word. Further, this negativity in response was larger
for related second words than for unrelated ones. They concluded that speech processing in
the presence of background noise changes activity of the cognition process such as auditory
working memory, attention, and semantic processing in anterior and central scalp areas.
It is assumed that improving the target signal while separated from the distractors spatially might at least partly be attributed to greater attentional focusing on the target, which is processed in the cognition areas, and which gives rise to different ERPs.

2.5.6. Rationale for recording the ABR, FFR and CAEP in this project.

To summarise, it was decided to record the following waveforms in this project for the following reasons:

- **ABR wave V (Chapter 4):** Due to the contrasts between the neural information coming up from the ipsi- and contralateral auditory pathways at the LL and the IC level, the processing of all spatial cues and their combination in the IC, and its contribution in generating the ABR wave V, it is reasonable to consider the recording of this wave to obtain information about the neural structures underlying SRM. In addition, ABR wave V is a common component in clinical applications and relatively easy to record.

- **FFR (Chapters 3 and 4):** As a steady-state and phased-locked response, FFR can provide valuable information about the extent that neural populations in the brainstem resist masking while encoding target stimuli. It might therefore be a suitable test for investigating the processing of spatial cues at the level of brainstem.

- **CAEP (Chapter 4):** Detecting spatial release from masking in the brainstem and then tracking it in the cortex might inform us about whether processing to achieve spatial release from masking occurs at both levels, under either passive or active listening conditions.

2.6. Summary

This chapter introduced CAP, (C)APD and spatial processing concepts that have been considered in a significant amount of research during the past decades. One type of (C)APD is known as spatial processing disorder – a reduced ability to identify the target sound in the presence of distractors, which is identified by the degree of SRM being smaller than normal.
To understand this disorder, first it is needed to identify how normal spatial processing occurs in the auditory system.

The auditory nerve is the first stage of enhancing the SNR of the target by shifting the dynamic range of its fibres to higher levels but only for moderate intensities. Once the auditory nerve conveys the neural information towards the cochlear nuclei, first stage of higher level processing will be applied based on neural characteristics. The units in the cochlear nuclei increase the SNR again. Then, the processed information is delivered to the next station via two paths. The DCN cochlear nucleus (path one) projects spectral information including spatial cues in the elevation plane monaurally to the IC units. The VCN cochlear nuclei (path two) send their processed information binaurally to the SOC, the MSO and LSO, to encode ITD and ILD spatial cues. These spatial cues, along with other signal features, will proceed to the IC via the LL. In the IC, all processed information, both monaurally and binaurally, converge into individual neurons. These neurons are assumed to behave as the auditory hub which represents the spatial location of a sound source and which sends this information toward the MGB. The MGB plays a role in enhancing the spectral cues and might refine other features of the signal. It then sends the processed information to the auditory cortex. The auditory cortex combines the incoming neural information including monaural and binaural cues together to determine the spatial location of the auditory signal in the background noise and to recognise the content of the transmitted massage.

To examine how the auditory system processes binaural/spatial cues objectively, AEPs need to be measured. AEPs are the specific neural activity arising from acoustic stimulation, which can be recorded within a half a second timeframe by placing several recording electrodes on specific locations of the scalp. To visualise these tiny electrical activities and to measure them in the presence of background noise of either neural and/or other sources, it is required to consider several factors, like stimulus repetition, intensity and length of the acoustic stimuli, using appropriate signal averaging techniques, amount of amplification, and selected filters. Based on neural response latency, three types of evoked responses can be classified: the ECochG, ABR & FFR known as early responses (the first 10 ms), the MLR or middle response (10-80 ms) and the CAEP or late response (80-500 ms).

To the author’s knowledge, the impact of SRM on FFR F0 amplitude had not yet been examined when target and distractors are in the same and different virtual spatial locations.
To address the thesis objective, ABR, FFR, and CAEP will be recorded utilising spatial virtual stimuli, and their relevant experiments will be introduced in the following chapters. In the next chapter, the first study will be presented. This study will investigate whether the FFR to target stimuli co-located with distractors at 0 degrees azimuth and then separated from the distractors by moving the distractors to the sides (±90 degrees azimuth) can be used to detect SRM. Furthermore, a factor of attention that might play a role in SRM is considered.
Chapter 3. Spatial release from masking (SRM): Human frequency-following response (FFR) correlation
3.1. Preface

In this chapter, one of the hypotheses of the thesis is explored, which is whether electrophysiological tests can be used to detect spatial release from masking (SRM). In addition, it is also examined whether attention modulates SRM. It is assumed that a significant effect of separation could be measured through the amplitude of the FFR and attention could modulate the FFR amplitude and SRM.

3.2. Introduction

Spatial release from masking (SRM) is the ability of the central auditory nervous system (CANS) to process target and distractor sounds depending on the change of their locations in space relative to each other. Changing the spatial location of target and distractor resulting changing in their spectral, temporal, and intensity cues which helps the CANS to select relevant information, i.e., the target signal, from coinciding but irrelevant information, the distractors (Alain, 2007, Alain and Arnott, 2000). It also refers to the improvement in the reception threshold of a signal when the target signal is spatially separated from competing sounds (Bronkhorst, 2000, Schneider et al., 2007). This process involves two mechanisms of central auditory processing: bottom-up and top-down. Bottom-up processing is based on interaural time and intensity differences (IIDs & ITDs) (see Sections 2.4.2.2.1 and 2.4.2.2.2) of incoming sounds (Bamiou, 2007, Shinn-Cunningham et al., 2005a) and top-down processing is based on selective spatial attention to the target sound (Huang et al., 2008, Huang et al., 2009, Rakerd et al., 2006).

During bottom-up processing, SRM presumably starts from the superior olivary complex (SOC) and progresses towards the inferior colliculus (IC) following the processing of target and distractors in the CN and the auditory nerve. The SOC contains two relevant structures: the medial superior olive (MSO) and the lateral superior olive (LSO). The majority of the neural fibres in the MSO and LSO are excited by stimuli arriving at both ears and based on their characteristics they are sensitive to ITD (Bazwinsky et al., 2003, Goldberg and Brown, 1969, Guinan et al., 1972b, Yin and Chan, 1990) and ILD (Boudreau and Tsuchitani, 1968, Tollin and Yin, 2002a) respectively (see Sections 2.4.2.2.1 and 2.4.2.2.2). At the level of the IC, the majority of the neural fibres are inhibited by stimuli at the ipsilateral ear and
excited by stimuli at the contralateral ear (Merchan et al., 2005, Oliver et al., 1994, Winer et al., 1996) (see Section 2.4.2.4). A small population of neural fibres in the IC are sensitive to ITD and are excited only by bilateral stimuli (Kelly et al., 1991). This contrast-based sensory processing of target and distractors from inferior neural regions towards the IC and at the level of IC, cause the target to stand out amidst the distractors. In fact, during the bottom-up processing of target and distractors, the target signal processing is driven by spatial cues (ILDs and ITDs), provide those for the target are different from those for the distractors (Du et al., 2009b). These differences in spatial cues help us to orient our attention towards the target signal, which then becomes more salient compared to the distractors. Further, the physical characteristics of the target stimuli may also cause the target to stand out more in comparison to the distractors (Schneider et al., 2007).

Top-down processing is a modulatory effect of the higher-order controlling mechanisms (Reynolds and Chelazzi, 2004). It modulates the activities of subcortical neural structures while these structures are processing and transferring target stimuli towards the higher centres. In fact, top-down processing generates a neural contrast that enhances (focuses attention on) the activity of bottom-up sensory mechanisms both for target as intended signals, and suppresses distractors as unintended signals. It relies on neural inputs from an executive control region encompassing prefrontal and parietal cortices (Ahveninen et al., 2006, Fan and Posner, 2004, Gazzaley and D'Esposito, 2007). Based on visual research, top-down processing modulates bottom-up processing via suppressing the “pop-out” characteristics of distractors (Beck and Kastner, 2009). This processing in the visual modality could also be extrapolated towards other modalities such as the auditory modality. Further, directing attention towards the specific location of a stimulus in space decreases the suppressive effect of the inhibitory neural networks. This inhibitory action is the result of top-down biasing to favour the attended location. It can be traced as an increased activity in the visual cortex (O'Craven et al., 1999). It is expected that focusing on a target auditory signal increases its representation at different stages of processing in the auditory pathways towards the cortex as well. This increasing in representation could potentially be based on two strategies: first, via the suppression of the distractors due to the dissimilarity between target and distractors; and second, via the improvement of the target signal by attending to its relative location in space and ignoring the distractors.
Electrophysiological measurements could provide suitable information for understanding neural mechanisms underpinning SRM. The accurate temporal resolution of these tools could help us examine bottom-up mechanisms of the auditory system while processing target signal in the presence of spatially separated distractors. Subcortical electrophysiological measurements that have been used with stimuli containing spatial cues are onset auditory brainstem responses (ABRs) (see Section 2.5.2.2) and sustained frequency following responses (FFRs) (see Section 2.5.3.2). ABRs and FFRs are two types of potentials that are generated predominantly in the brainstem when evoked by appropriate acoustical stimulation. The ABR is a transient electrical activity, aroused by brief stimuli such as tone bursts or clicks and based on synchronous activity of the eighth cranial nerve fibres to the IC and thalamus (Hall, 2007, Jewett et al., 1970) (See Section 2.5.2.1). The FFR, in contrast, is an ongoing neural activity that is phase-locked to the individual cycle of a periodic low-to-middle frequency stimulus and originates from about the same regions as the ABR at the brainstem level (Ballachanda et al., 1994, Bledsoe and Moushegian, 1980, Hoormann et al., 1992, Moushegian et al., 1973, Worden and Marsh, 1968, Marsh and Worden, 1968). The FFR reflects the fundamental-frequency (F0) characteristic of a complex signal such as speech and some higher harmonics (e.g., Krishnan et al., 2004, Greenberg et al., 1987, Russo et al., 2004). The FFR to F0 of a target stimulus is considerably resistant to interference of distractors even at lower SNRs (Cunningham et al., 2002, Li and Jeng, 2011, Russo et al., 2004, Song et al., 2011). Because of these specific properties of the FFR, this electrophysiological measure could be suitable for the investigation of the neural mechanisms underpinning SRM during bottom-up processing of a target signal in a noisy environment.

To our knowledge, only two studies have considered distractors while presenting target stimuli using FFR in animals. In animal binaural unmasking studies (e.g., Du et al., 2009a, Du et al., 2009b), introducing an ITD disparity between the target and masker significantly boosted FFR amplitudes. Du et al. (2009b) demonstrated that FFRs originating from the IC while evoked by target signals in the presence of distractors in rats were enhanced via presenting an ITD between the target signals and distractors. Further, FFR recording from the lateral nucleus of the rat’s amygdala demonstrated that in the presence of a correlated noise interaurally when ipsilateral babble (or tone complexes) leads contralateral babble with respect to ITD, binaural unmasking occurred similarly (Du et al.,
Brainstem FFR is also unmasked in humans by shifting either signal or noise by $\pi$ radians, referred to as binaural masking level difference (BMLD) (Wilson and Krishnan, 2005). In other studies that have been conducted on FFR amplitude, the spatial cues of target stimuli have been manipulated without introducing any distractors (Ballachanda and Moushegian, 2000, Clark et al., 1997, Krishnan and McDaniel, 1998) (see Section 2.5.3.2). The findings of these studies were that binaural FFR amplitude decreased with increasing ILD and ITD. They concluded that lower brainstem regions are the initial sites of binaural processing reflecting spatial cues. Taken together, these studies demonstrate that at least some of the processing of ILD and ITD cues occurs at the level of the brainstem, and that FFR could be a potential tool for representing these spatial cues. They also demonstrated that individual or combined spatial cues in a lateralisation paradigm and in a quiet condition could be reflected using the FFR. Furthermore, the FFR was recorded using target stimuli in the presence of background noise, but only by using individual spatial cues. Although these studies evaluated the two major cues in binaural processing and showed that this tool has the potential to demonstrate normal and therefore possibly abnormal binaural processing at the brainstem, they are not well positioned to provide a complete picture of spatial processing. They did not show that these same cues could be used to separate a target stimulus from spatially separated competing stimuli. It is assumed that the FFR can represent SRM in a more naturalistic scenario (with background noise in the form of for example babble) in which both spatial cues are investigated simultaneously. This is what it is tried to explore in this experiment.

Whether the cognitive resources underlying attention could modulate the activities of subcortical structures have not yet been fully investigated. At least at the cortical level studies in humans have demonstrated that attention has a strong effect on neural activities induced by acoustic stimuli. Attention to the target stimuli generally results in larger amplitudes and shorter latencies of P1, N1 and P2 in cortical responses (Hansen and Hillyard, 1980, Hillyard et al., 1973, Näätänen et al., 1992, Schwent and Hillyard, 1975, Woldorff et al., 1987, Zhang et al., 2014). However, at the subcortical level, investigations have not yet been able to detect an effect of attention on the brainstem by means of ABR (Näätänen et al., 1992, Picton and Hillyard, 1974). On the other hand, FFRs evoked by pure
tones, missing-fundamental stimuli, and dichotic vowels seem to have demonstrated the effect of attention on both amplitude and latency (Galbraith and Arroyo, 1993, Galbraith and Doan, 1995, Galbraith et al., 2003, Hoormann et al., 2004, Galbraith et al., 1998). Galbraith et al. (1998) reported that amplitudes and latencies of FFR F0 were larger and shorter, respectively, when vowels were attended than when ignored. Using pure tone stimuli while detecting frequency target stimuli in either an auditory or visual detection task, Galbraith et al. (2003) found that overall FFR amplitudes significantly increased during the auditory attention task. Effects of attention have also been demonstrated on FFR amplitude using missing fundamental frequency and using pure tone stimuli with varying in their durations and intensities while being attended than ignored (Galbraith and Doan, 1995). Although the effect of attention on the FFR has been investigate in some studies, these studies were nor using specific target stimuli or distractors. As a result, the effect of top-down attentional mechanisms on bottom-up sensory driven mechanisms in the presence of distractors needs to be established yet. It is probable that attentional top-down mechanisms modulate the sensory bottom-up binaural processing as individuals selectively attend to target speech while ignoring distracting sources. Hence, it would be interesting to know the effect of attending to the target on objective measures of spatial processing. Suga and Ma (2003) reported that the characteristics of the IC responses have been modulated by stimulation of cortical neurons. Rinne et al. (2008) ascertained using fMRI that activation of the human IC was significantly modulated by auditory attending tasks and this activation was closely dependent on the spatial direction of the attention.

The two main objectives of this study are to examine whether

- the phase-locked neural activity in the brainstem, as measured by the FFR, reflects SRM;
- this neural representation of SRM is modulated by attention.

In particular, it will be examined whether there is an effect of attention on the FFR, and an interaction between SRM and attention. It is hypothesised that attention-invoked top-down influence could modulate SRM as reflected in the FFR. That is, if processing to achieve SRM occurs in the brainstem but is under executive control, such an interaction might be visible in the amplitude of the FFR. If the objective results reflect behavioural SRM,
it may be possible to use objective tests to detect spatial processing and its disorder (SPD) at an early age.

A speech masker can effectively mask speech stimuli depending on the number of talkers incorporated in the masker. It has been reported that when the speech masker incorporates two talkers, informational masking is high (Freyman et al., 2004). To meet the aims of this experiment, hence speech elements as target stimuli in the presence of ongoing, same-talker, two-talker babble as distractors will be introduced.

To identify whether SRM is modulated by an attentional effect of top-down mechanisms at the level of the brainstem, a factor of attention, achieved by using both passive and active listening conditions, will be considered. The motivation to consider this condition in our experiment was that the neural substrate of the top-down mechanism at the level of brainstem is poorly understood, despite the large body of investigations on attentional process at the level of the cortex.

### 3.3. Materials and methods

#### 3.3.1. Participants

Nineteen normally hearing adults (9 males and 10 females) with a mean age of 32.2 (standard deviation of 6.3 years, ranging from 20 to 42 years) participated in the study. The participants were selected from list of volunteers at the National Acoustic Laboratories (NAL) and students from the Macquarie University via public announcement, hence unbiased non-probability sampling. Hearing sensitivity in all participants was 20 dB HL or better at octave frequencies from 125-4000 Hz with symmetric thresholds bilaterally. All participants provided informed consent prior to testing and were compensated for their participation. This study was approved by the Australian Hearing Human Research Ethics Committee.

#### 3.3.2. Procedure

Participants were seated in a comfortable recliner in a dimly lit, double-walled and electromagnetically shielded sound-treated booth. Each session consisted of 12 runs, with two attention (attend and ignore), two spatial (co-located and separated), and 3 SNR (-5, 0
and 5 dB SNR) conditions. During the attention condition, subjects were asked to focus on the target stimuli. Deviants were presented randomly among the target stimuli. Participants had to respond to these deviants by pushing a button. If the reported number of button presses was more than 10% different from the actual number of deviants in the sequence, indicating the participant was losing attention, the sequence was repeated. In total, 12 out of 228 sequences (12 sequences times 19 subjects) were repeated. On the other hand, during the non-attention condition, subjects were advised to ignore any stimuli and encouraged to be relaxed as much as possible and fall asleep if preferred. Test sequences were randomised and counterbalanced across subjects. Altogether, the required time for completion of the experiment was around 2 hours.

3.3.3. Stimuli

A steady-state target vowel /u/ was synthesised using Praat software (Boersma, 2001). The fundamental frequency F0 was 110 Hz, representing a male voice. First and second formants were chosen at 310 and 870 Hz respectively. The vowel /u/ was chosen because its fundamental frequency and formants are the lowest frequencies of all the vowels, and is therefore expected to elicit the largest amplitude FFRs. The overall duration was 250 ms. The presentation level was 60 dB SPL. The deviant vowel /u/ was exactly the same concerning physical characteristics as the target stimulus but 8 dB lower in presentation level.

The target vowel, as shown in the top panel of Figure 21, was convolved with head-related transfer functions (HRTFs) and positioned at 0 degrees azimuth for all recordings. HRTFs are series of measures that reflect the temporospectral characteristics of a sound wave as it travels from a particular location in space to the eardrum. The HRTFs describe the sound pressure at the eardrum divided by the sound pressure in the undisturbed sound field, and is most important for high-frequency sounds because their wavelengths are close to the size of head, pinna and torso (Blauert, 1997, Butler, 1969, Klumpp and Eady, 1956). Therefore, a pair of HRTFs for both ears can be developed and used to synthesise any particular sound location spatially for virtual delivery under headphones. When the sound is presented to listeners under headphones, a realistic perception of the sound location in a spatial location could be simulated (Wenzel et al., 1993). Although calculating individual HRTFs for each subject might be ideal, they may also behave as a confounding factor.
because of different HRTFs in different individuals. Alternatively, non-individualised HRTFs have been developed in an anechoic chamber using the Knowles Electronics Manikin for Acoustic Research (KEMAR). KEMAR’s ear canal and eardrum were simulated using a Zwischen coupler fitted with a half-inch microphone. The HRTFs were created using sine waves ranging in frequency from 0.05-20 kHz which were delivered through a loudspeaker at a one-meter distance referenced to the centre point of KEMAR’s head. KEMAR was positioned in the situations to make recordings at 0°, +90°, and -90° azimuth. The elevation angle was 0° for all recordings made. The amplified signal created by the microphone at the eardrum of KEMAR was inserted to one channel of the signal analyser, and the input to the loudspeaker was inserted to the second channel. Then, the transfer function was calculated as the ratio of the microphone output spectrum to the loudspeaker input spectrum. The HRTFs then were convolved with both target and distractor wave files using MATLAB (MathWorks Inc, 2012). In the current study, the synthesised /u/ vowel was first converted to the frequency domain by a fast Fourier transform (FFT), multiplied by non-individual HRTFs followed by a conversion back into the time domain for delivery through headphones.

Two maskers were introduced in the form of two discourses narrated by the same male speaker (David Wigney) from the NAL Speech and Noise for Hearing Aid Evaluation CD (disk 1, track 3, left channel). These discourses were recorded utilising the procedure similar to that has been used in the International Long-Term Average Speech Spectrum (ILTASS) (Byrne et al., 1994) and were sampled at the rate of 44,100 Hz. During matching of the discourses, the relative intensities of each 1/3rd octave frequency of the signals matched the average relative intensities of speech produced by male and female talkers (Keidser et al., 2002). These discourses were developed for previous studies conducted at NAL.

Figure 21 (mid and bottom panel) demonstrates the waveforms, the spectrograms, and the spectral information of the target and distractor stimuli that were utilised in the current experiment. The maskers were presented simultaneously during target presentation, which was located at 0 degrees azimuth. Two spatial conditions were considered: a co-located and a separated condition, introduced by convolving the maskers with non-individualized HRTFs. The HRTFs at both 0° and ±90° azimuth have been recorded in an anechoic chamber using the Knowles Electronics Manikin for Acoustic Research (KEMAR) similar to the procedure for the target stimulus. In the co-located condition, both maskers
were placed in the front (0 degree azimuth). In the separated condition, one discourse was placed at -90 degrees and the other at +90 degrees azimuth.

Figure 21. The spectrogram, and spectra of /u/ target stimulus (top panel) combined with two-talker discourses for two channels in co-located condition (middle panel), and separated condition (bottom panel).
Stimuli were delivered using MATLAB in combination with a high fidelity multichannel RME soundcard (ADI-8 QS), through electromagnetically shielded Etymotic Research ER-3A insert earphones connected with 28 cm air tubes. Stimuli were calibrated in a HA2 2-cc coupler, 1 inch microphone (Brüel&Kjær (B&K) 4144), 1 to ½ inch adaptor (B&K 0374), a microphone preamplifier (B&K 2660) and a measuring amplifier with scale number SA 0262 (B&K 2636) which used a 30-second averaging time window to avoid large fluctuations in amplitude. The sound files were stored in the computer with MATLAB code for controlling the presentation of the file waves. The computer output was input to B&K 2636 sound measurement amplifier. The broadband signal was filtered from 22.4 Hz to 22.4 kHz. The measurement amplifier calculated the root mean square (rms) envelope of the input signal using an averaging time of 1 second and was fixed at 60 dB SPL.

To provide a variation that the participants would be required to detect in the attended condition, approximately 5% of the target stimuli were dropped in presentation level by 8 dB (Galbraith and Arroyo, 1993, Galbraith and Doan, 1995) and randomly distributed throughout the target stimuli.

As it is hypothesised that different SNRs may influence spatial processing in different ways, tests were conducted at three global SNRs: -5, 0, and 5 dB, defined as the ratio of the target level averaged across its 250 ms duration, relative to the level of a single masker, averaged across its entire duration (which was several minutes) in the co-located condition. The SNRs were chosen to cover a range of SNRs that would provide varying degrees of difficulty in differentiating the target from the noise, but within a range that is realistic for conversations. When both noise maskers were presented simultaneously, the SNR dropped on average to -3 dB. In addition, because of the HRTFs, the objective SNR of the target was 0.9 dB poorer in the separated condition when compared to the co-located condition. This reduction in target level guaranteed that any increases found in the target FFR amplitude when separating maskers from the target was not just caused by an increase in SNR.

During data analysis, it became clear that the variability of the masker level, and hence the SNR of the target stimulus, was influencing the results. This analysis was based on both statistical analysis of FFR SNR and the distribution of signal and noise for the entire length of an individual condition. To reduce the SNR range per condition, ‘local’ stimulus SNRs were defined and specific ranges (low, medium, and high local SNRs) investigated separately, as defined in Figure 22, ‘Local’, or short-term SNR, was calculated from target
and masker powers taken during target presentation corresponding to each 250 ms time frame over which the target vowel occurred (Vondrasek and Pollak, 2005). Local SNRs are shown in relation to global SNRs in the top panel of Figure 22. The distributions were the same for all global SNRs, except for a shift of 5 or 10 dB. The three global SNRs were pooled together, and one distribution drawn, as depicted in the bottom half of Figure 22. The range of local SNRs was divided into three parts: a low, a medium and a high SNR range, with the number of data points equal in each range. The ranges were as follows (in dB SNR): -12.9 to -2.4 (low), -2.4 to 3.2 (medium), 3.2 to 60.4 (high).

![Graph showing distribution of local SNRs for different global SNRs](image)

**Figure 22.** Distribution of local SNRs of -5, 0, and 5 dB (top panel). Distribution of local SNRs for low, medium, and high ranges (bottom panel).

### 3.3.4. EEG recording

We considered three recording channels to identify whether different electrode montages might pick up neural information from different neural population loci by means of two vertical (forehead at the hairline referenced to the 7th cervical vertebra, and referenced to both mastoids) and one horizontal (the left ear lobe referenced to the right ear lobe) montages. Galbraith et al. (1998), using these two electrode montage directions,
found that the vertical channel better conveys F₀ and the horizontal channel better identifies high frequencies components of /e/ and /a/ vowel stimuli. Recording of high frequency components of stimuli is also enhanced while using high forehead as an active electrode linked to mastoids. Further, applying high forehead as an active electrode linked to the 7th cervical vertebra shifts the recording of neural populations underlying FFR towards more rostral generators (Krishnan, 2007). In the case of the auditory brainstem response, Picton et al. (1991) and Scherg and Von Cramon (1985) have suggested that the horizontal array represents earlier and the vertical array represent later components of the ABR. Therefore, it is reasonable to record the FFR in different channels as these channels might provide supplementary information while investigating attention and SRM.

Hence, electroencephalographic (EEG) activity was recorded via gold-plated cup electrodes in three channels according to the 10-20 electrode placement system (Jasper, 1958): channel one with the active electrode on the centre of the high forehead near the hairline (Fp) referenced to the 7th cervical vertebra (C₇) (a protuberant small area on the back of the neck that with tilting the head downward can be palpable), channel two with the active electrode on left ear lobe referenced to right ear lobe (A₁ and A₂) (known as a horizontal montage), and channel three with the active electrode of channel one referenced to linked right and left mastoids (vertical montage). The ground electrode was located at the midline lower forehead (Fpz) (above the nasion). The interelectrode impedances were maintained below 1 kΩ and were measured pre-, peri- and post-sessions. The EEG signals were recorded using an Intelligent Hearing System (IHS) with a 150,000 amplification gain while analogue filtered with a frequency bandpass of 50-3000 Hz (6 dB/octave roll-off) and 50 Hz notch filter. The EEG sampling rate was 10 kHz. The start of each stimulus was recorded precisely in the digitally recorded EEG data stream (within 25 µs). Per condition, 1100 stimuli were presented with a 350 ms stimulus onset asynchrony (SOA). Hence, the repetition rate of stimuli was 2.85/sec. The EEG analysis window was 275 ms commencing at the stimulus onset.

3.3.5. Data analysis

We recorded 1100 epochs of the target stimuli, not including deviant epochs, as these were discarded in each recording condition. For each recording, the 100 epochs with
the largest absolute values across each epoch were artefact-rejected off-line. The remaining 1000 epochs were combined using noise-power weighted averaging. The noise power of each epoch was calculated by taking the sum of the squares of all samples in the epoch. Then, all epochs were weighted by the inverse of their noise power, their sum taken, and normalised such that the sum of all inverse noise powers equalled 1 (Hoke et al., 1984). An FFT was applied, with a bin width of 5 Hz. This bin width seemed optimal as it approximated the inverse of the stimulus duration, and it seemed to be the narrowest bin width that would capture most of the energy of each FFR component. Amplitudes of the different FFR components (F0, its harmonics up to the 5th harmonic, F1 and F2) were determined. Noise amplitudes around the signal bins were calculated through the root-mean-square (RMS) amplitude of 10 neighbouring non-signal bins on each side of the signal bin. Based on signal and noise power values, a response SNR was determined for each signal component.

3.3.6. Statistical analysis

To identify the strength of recorded signals relative to the recorded noise, a preliminary examination of the response SNRs for each FFR component was carried out, which prompted the focus of analysis on the F0 FFR component. The mean SNRs of the higher harmonics (2F0 to 5F0) and formants F1 and F2 ranged between -3 and 2 dB (Figure 23). The majority of these responses were not found to be significantly different from the noise floor, by comparing the F-ratio formed from the response power and the average power of the surrounding noise bins (Dobie and Wilson, 1996). Hence, it was decided to leave the higher harmonics and formants out of the analysis. In addition, all results, except Section 3.5.5 investigating the magnitude of effects, relate to the amplitude of the fundamental frequency F0 of FFR. Moreover, the analysis indicated that channels one and three were similar in SNR (p = 0.052), and had significantly higher response SNRs than channel two (the horizontal montage; p<0.00001) (Figure 24). Therefore, all analyses were based on data collected through channels one and three. During this preliminary analysis, it is also completed a normality test to examine whether our data met the requirement for a parametric analysis. First, the frequency distribution was visually inspected via normal plots (histogram and boxplot) and second by performing the Kolmogorov-Smirnov (K-S) normality test with a significance level of 0.05. As the means of data in all conditions accurately
represented the centre distribution of the data and given that the sample size was large enough, the requirement for a parametric test was considered to be met.

After preliminary statistical analysis, data from one participant with an extreme amplitude (greater than 5 SD units from the mean) were excluded from further analysis. Thus, eighteen out of nineteen participants were included for further analysis.

Figure 23. Means and standard errors of response SNR for each component of the FFR, averaged across recording channels.

Figure 24. Means and standard errors of response SNR of FFR in recording channels, averaged across response components.
Analysis of data based on local SNRs returned larger differences between test conditions than global SNRs, presumably because of a reduced variability of masker level and hence stimulus SNR and response amplitude. Therefore, all the following main results and figures are presented based on local SNRs, but are accompanied by parallel results for the global SNRs. As will be shown, both global and local SNR data are consistent with each other, which add to the confidence of the results obtained in this study.

For the analysis of the data, a 4-way 2x3x2x3 repeated measures analysis of variance (ANOVA) was applied to the F0 amplitude with spatial location (co-located versus separated), SNR (-5, 0, 5 dB), attention condition (attention versus non-attention), and channel (channel one Fz-C7, and channel three Fz- A1 and A2 (both mastoids) as repeated-measures variables.

### 3.4. Results

This section first will present the main effects of SRM, stimulus SNR, recording channel, and level of attention on the amplitude of the fundamental frequency component F0 of FFR. The interactions between these measures will then be presented. To summarise the results obtained in this chapter, it is calculated the magnitude of the SRM in terms of amount (in dB) by which the input SNR would need to be changed to achieve the same response amplitude that spatial separation provides. Expressing the results in this way will make it possible to compare the results obtained in this chapter with the results obtained in the second study, described in Chapter 4.

Statistical analysis showed significant effects of attention, SNR, channel, and a significant interaction effect of SRM with SNR (Table 1). None of the 3-way and 4-way interactions was found to be significant.

### 3.4.1. Evidence for SRM in FFR F0 amplitude

Figure 25 and shows FFR amplitudes for spatially co-located and separated conditions at different SNRs. No significant main effect of SRM was found on FFR amplitude (Table 1).
Figure 25. Means and standard errors of FFR F0 amplitude in spatially co-located and separated conditions as a function of stimulus SNR (low, medium, and high), averaged across the two attention conditions and channels one and three.

The lack of a main effect of SRM on F0 amplitude was confirmed when analysing the global SNR data as well ($F_{(1, 17)}=0.76$, $p=0.40$).

Table 1. Summary of the statistical results of a 4-way 2x3x2x2 repeated measures ANOVA with SRM, SNR, channel and attention condition as repeated-measures variables.

<table>
<thead>
<tr>
<th>ANOVA</th>
<th>$F$-value</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRM</td>
<td>$F_{(1, 17)}=0.16$</td>
<td>$p=0.70$</td>
</tr>
<tr>
<td>Attention</td>
<td>$F_{(1, 17)}=8.96$</td>
<td>$p=0.008$</td>
</tr>
<tr>
<td>SNR</td>
<td>$F_{(2, 34)}=11.29$</td>
<td>$p=0.00017$</td>
</tr>
<tr>
<td>Channel</td>
<td>$F_{(1, 17)}=22.90$</td>
<td>$p=0.00017$</td>
</tr>
<tr>
<td>SRM x attention</td>
<td>$F_{(1, 17)}=0.03$</td>
<td>$p=0.86$</td>
</tr>
<tr>
<td>SRM x SNR</td>
<td>$F_{(2, 34)}=5.20$</td>
<td>$p=0.011$</td>
</tr>
<tr>
<td>Attention x SNR</td>
<td>$F_{(2, 34)}=1.99$</td>
<td>$p=0.15$</td>
</tr>
<tr>
<td>SRM x Channel</td>
<td>$F_{(1, 17)}=2.78$</td>
<td>$p=0.11$</td>
</tr>
<tr>
<td>Attention x Channel</td>
<td>$F_{(1, 17)}=4.28$</td>
<td>$p=0.054$</td>
</tr>
<tr>
<td>SNR x Channel</td>
<td>$F_{(2, 34)}=3.23$</td>
<td>$p=0.052$</td>
</tr>
</tbody>
</table>
3.4.2. Effects of Stimulus SNR on FFR F0 amplitude

3.4.2.1. Main effects of stimulus SNR

The effect of stimulus SNR on F0 amplitude is shown in Figure 25 and Figure 26. The latter figure illustrates the amplitude spectra per stimulus SNR, grand averaged over all 18 participants. With increasing stimulus SNR, the mean F0 amplitude became significantly larger (Table 1).

The findings were confirmed when analysing the global SNR data as well. Mean F0 amplitudes increased significantly ($F_{(2, 34)}=37.69, p<0.00001$) with increasing global stimulus SNRs (-5, 0, and 5 dB).

Figure 26. Grand averages of 18 participants showing the FFR in the frequency domain for low, medium and high SNRs, collapsed across 2 attention and 2 spatial conditions, and channels 1 and 3.

3.4.2.2. Interaction effects of SRM and SNR

The interaction effect between SRM and stimulus SNR was significant (Figure 25, Figure 27 and Table 1). As the figure demonstrates, SRM occurs only at the lower SNR.

Post hoc comparisons using a Tukey’s honest significant difference (HSD) test indicated that the mean F0 amplitude for the low SNRs in the co-located condition ($M=43.24\text{nV}, SE=3.61$) was significantly smaller ($p=0.03$) than the separated condition ($M=53.12\text{nV}, SE=4.05$). There was no significant difference ($p=0.77$) for medium SNRs between co-located ($M=57.48\text{nV}, SE=3.72$) and separated ($M=56.39\text{nV}, SE=4.73$) conditions. Nor was there a significant difference for high SNRs ($p=0.33$) between co-located ($M=65.05, SE=4.67$) and separated ($M=60.19, SE=4.38$) conditions.
Figure 27. Grand averages of FFR F0 amplitude over 18 participants of the response spectra for stimulus local SNRs: low, medium and high in co-located and separated spatial conditions, averaged across two attention conditions and channels (one and three).

Taken together, these results suggest that spatially separating the distractors from the target does have an effect on F0 amplitude, but the effect depends on stimulus SNR. Specifically, our results suggest that when the listening environment is noisiest, spatial separation has the greatest effect.

A similar effect of SRM was found at -5 dB SNR in global SNRs (p=0.056), akin to the results for the low local SNR condition.

3.4.3. Effects of attention on FFR F0 amplitude

3.4.3.1. Main effects of attention

A significant main effect of attention on the FFR F0 amplitude was found (Table 1 and Figure 28). Attending to the target stimulus results in larger FFR F0 amplitudes than ignoring the target stimulus. This result is consistent with results from FFR experiments that have considered attention (Caird and Klinke, 1983, Fuzessery et al., 1985, Galbraith and Arroyo, 1993, Galbraith et al., 1998, Galbraith and Doan, 1995, Galbraith et al., 2003).

For global SNRs, a statistically significant effect of attention was found as well ($F_{11, 177}=7.42, p=0.01$).
3.4.3.2. Interaction effects of attention and SRM

No significant interactions effects were found between SRM and attention (Table 1 and Figure 29). Analysis of global SNRs could not reveal any significant interaction effect between attention and SRM as well ($F_{1, 17} = 0.03$, $p = 0.86$).
3.4.3.3. Interaction effects of attention and SNR

No significant interaction was found between attention and stimulus SNR (Table 1 and Figure 31). This was confirmed by the analysis of the global SNRs ($F_{(2, 34)} = 1.25, p=0.3$). However, Figure 30 and Figure 31 show that there is a non-significant trend for the effect of attention to be greater at lower SNRs. This potentially could justify additional research into the effect of attention on SRM at low SNRs.

Figure 30. Grand averages of FFR F0 amplitude over 18 participants of the response spectra for stimulus local SNRs: low, medium and high in attention and non-attention conditions, averaged across two spatial conditions and channels (one and three).

Figure 31. Means and standard errors of F0 amplitude in attention conditions as a function of stimulus SNR, averaged across spatial conditions and channels one and three.
3.4.4. Effect of channel on evaluated parameters

Figure 32 illustrates the effect of channels on FFR F0 amplitude. A significant main effect of channel was found (Table 1). Channel one (Fp-C7) has a larger F0 amplitude than channel three (Fp- both mastoids).

Although amplitudes in channel one are significantly larger (p = 0.0098) than those in channel three, SNRs are not significantly different in both channels (p = 0.052) (Figure 24) due to more noise in channel three. Given that, when averaged across all inputs SNRs, SRM is not significant when averaged across channels, it is do not expected that SRM would be significant in either channel individually. This is supported by the non-significant interaction effect between SRM and channel (Table 1).

3.4.5. Magnitude of effects

The amount of increase in neural response associated with SRM, SNR or attention has so far been expressed as a change in response voltage. However, it is more informative to know by how much the stimulus SNR would have to improve to achieve the same
increase in response magnitude. This makes comparison with behavioural studies and the results of the study in the next chapter easier.

When collapsing across attention and channels, the following FFR F0 amplitudes (in nV) were obtained in spatial co-location as a reference condition: 39.79, 56.98, and 69.70 nV for -5, 0, and +5 dB global SNR respectively. Hence, the increase of the FFR F0 amplitude in nanovolt per decibel is 2.99 nV/dB, based on an increase of 29.91 nV for a 10 dB change in SNR.

When collapsing over all attention conditions and both channels for local stimulus SNRs, as depicted in Figure 25, the F0 amplitude increase due to SRM for low, medium, and high local SNRs is 9.88, -1.11, and -4.86 nV. These amplitudes correspond to an effective 3.30, -0.37, and -1.63 dB improvement of the stimulus SNR while keeping other stimulus conditions constant.

3.5. Discussion

When speech and distracting sources are spatially separated in a noisy environment, speech recognition improves appreciably. This improvement in speech recognition relies on the availability of specific cues such as ILD, ITD and spectral information that are processed during the progress of signal and noise through the central auditory pathways. The central auditory pathways in the present study were investigated to determine whether the FFRs evoked during binaural stimulation can demonstrate SRM objectively, and whether and how a change of the target SNR and mode of attention might affect this process. The findings of this study suggest that separating distractors spatially from the target results in significant enhancement of FFR F0 amplitude, but only when the target is presented at sufficiently low SNRs. With increasing noise and hence reducing SNR, the necessity to use spatial information to separate a target from competing sounds is progressively increased. At lower SNRs, the decorrelation between competing sounds at the binaural inputs is greater when compared at higher SNRs, resulting in a greater release from masking. It is possible that SRM occurs at several levels within the auditory system. Moreover, in the present study the significant effects of attention and SRM were found only for FFR fundamental frequency (F0) across all conditions that have been tested. F0 conveys linguistic distinctions, such as segmental features; prosodic or suprasegmental features, such as intonation; paralinguistic
information, such as emotional state of the speaker; sex and age (McRoberts et al., 1995). Finding significant effect of SRM only for FFR F0 suggests that neural populations underlying the encoding of SRM process salient fundamental cues of the incoming target signal as these cues can also convey information of spatial cues. Moreover, finding significant effect of attention on FFR F0 suggests that top-down attentional mechanisms modulate neural population activities underlying the encoding of the salient fundamental cues to filter sensory information.

### 3.5.1. Binaural spatial release from masking at the brainstem level

This study showed for the first time that by considering both spatial (ITD and ILD) and spectral cues through applying non-individual HRTFs to both speech target and speech distractors, the amplitude of the FFR F0 increased when distractors are moved away from a co-located condition (at 0° azimuth) to a separated condition (at ±90° azimuth) (Figure 25). This demonstration of FFR F0 amplitude enhancement under spatial separation of distractors from the target suggests that the phase-locked neural activities the FFR originates from, may indeed reflect information regarding neural activities underpinning spatial release from masking (SRM). This result is generally consistent with the literature showing that a spatial dislocation between a target and distractors in binaural processing enhances auditory presentation of the target (Caird et al., 1991, Du et al., 2009a, Du et al., 2009b, Lane and Delgutte, 2005, Lin and Feng, 2003, Mandava et al., 1996, Palmer et al., 2000, Wilson and Krishnan, 2005). In particular, similar findings have been reported by Du et al. (2009b) who found that FFRs originating from the IC while evoked by target signals in the presence of distractors in rats were enhanced when there was an ITD between the target signals and distractors. The result obtained in the current experiment also is in agreement with Du et al. (2009a) study, which demonstrated binaural unmasking utilising FFR, although in their study they recorded the FFR from the lateral nucleus of the rat’s amygdala in the presence of noise in a lateralisation ITD paradigm. In addition, because the current experiment was performed using speech target and distractors stimuli in different spatial locations, it is expected that these stimuli elicited electrical activities of different neural populations that are responsible for encoding of ITD, ILD cues associated with spectral cues.
Similar SRM has been reported for binaural masking level difference (BMLD) which occurs by shifting either signal or noise by π radians (see Section 2.5.3.2) (Wilson and Krishnan, 2005). Though both the current study and those mentioned above have demonstrated the SRM phenomenon, it is likely that the underlying neural mechanisms in these studies might be different. In the Du et al. (2009a) study for example, SRM occurred at >2 kHz while there was only a disparity in time between target and distractors binaurally. The target was perceived in one ear and the distractors were either perceived in the other ear (using an ITD of 0.1 ms) or at the centre of the head (using ITD of 0.2 ms). This mode of presentation results in creating two objective acoustic images: one for the target and the other for the distractors. These two different images are probably processed by two different ITD-detector neural populations, each with sensitivity to encode incoming stimuli from different spatial locations. On the other hand, BMLD is more efficient below 1-2 kHz (e.g., Caird et al., 1991, Mandava et al., 1996, Wilson and Krishnan, 2005) and is based on the cross-correlation of target and distractors in two ears exerted by phase differences between the two ears. These differences in phase occur either by the interaural decorrelation of the target response to interaurally correlated distractors or vice versa. In either case it might be that different neural populations underlay different binaural configurations (Palmer et al., 1999). Nevertheless, some overlap in neural activities underlying ITD-based and BMLD-based unmasking could be assumed, though the earlier is based on time dislocation and the latter is based on phase dislocation between target and distractors. Finally, in the present study, it is tried to use more naturalistic acoustic stimuli, simulating a more natural acoustic environment. Examining the acoustic stimuli under this naturally simulated paradigm gives an insight into how the central auditory pathways behave when the locations of the distractor stimuli change in complex acoustic environments. This setup might either stimulate different neural structures in either the co-located spatial location, where both target and distractors were perceived in frontal location. Hence, these stimuli activated neural populations responsible for encoding ITD and ILD cues at 0⁰ azimuth. This condition, however, may partly have similarity with those conditions mentioned in the above studies (0 ms ITD and S0N0). Alternatively, the setup might stimulate in the separated spatial location, where the target was perceived in the frontal location and the distractors were perceived in lateral locations. Hence, they activated different neural populations responsible for encoding ITD and ILD cues at 0⁰ and 90⁰ azimuths respectively. This condition is different
from those stated in the above studies, hence in the current study the phase-locked neural activities of the neural elements underlying SRM that has generated the FFR is not similar to the FFRs having been recorded in the above-mentioned studies. More specifically, encoding of spatial cues including ILD (Tollin and Yin, 2002a) and ITD (Jeffress, 1948, Tollin and Yin, 2005, Yin and Chan, 1990) start at the lateral superior olivary complex (LSO) medial superior olivary complex (MSO), respectively. Spectral cues are believed to start at the dorsal cochlear nucleus (DCN). These neural arrangements are likewise evidenced by histological studies (Boudreau and Tsuchitani, 1968, Boudreau and Tsuchitani, 1970, Caird and Klinke, 1983, Guinan et al., 1972b, Joris and Yin, 1995, Tsuchitani, 1977). According to these neuroanatomical and histological studies, it is assumed in the current study that in both co-located and separated conditions, where both target and distractors are bilaterally presented simultaneously, the MSO fibres operate at maximum rate while the LSO fibres respond minimally to both target and distractors (Yin and Kuwada, 1983). In the spatially co-located condition, both target and distractors reach the auditory system at same spatial locations, creating acoustically fused images. The acoustic image of the target stimuli created in this condition is blurred due to destructive effects of distractors resulting in decrease of the FFR F0 amplitude. In the spatially separated condition, on the other hand, the target and the distractors reach the auditory system at different virtual locations hence at different times. The acoustic images of the target stimuli still fuse in the midline (mid sagittal plane) but are distinct from those images that are created under the co-located condition, suggesting that although both target and distractors are coming from the frontal location activating the same ILD- and ITD-neural detectors, separating distractors from the target enhances the FFR F0 amplitude. This advocates a spatiotopic organisation of the neural populations underpinning SRM at the level of the brainstem: under this condition, neural populations that are sensitive to frontal locations (through either ITD or ILD cues or both) have to respond to both the target and the distractors in the co-located condition. In the separated condition, however, they respond more weakly to the distractors, thus enabling more of their available firings to respond to the target. This increased responsiveness is evident as the stronger FFR in the separated condition.

Taking into account that (1) the IC is a potential generator of the FFR, that (2) the binaural and monaural pathways, which convey spatial and spectral cues respectively, combine at the level of single neural cells in this generator (Oliver and Huerta, 1992, Vater
et al., 1995), and (3) the significant improvement in the FFR amplitude in the separated condition in the present study, it seems highly probable that at least some SRM occurs at the level of the brainstem. This is supported by studies reporting that the contralateral dorsal nucleus of the lateral lemniscus (DNLL) decreases the IC FFR in a quiet condition. This is evidenced by the FFR increasing when the contralateral DNLL was disabled. In masking conditions, however, when the distractor was located at the ipsilateral side, disabling the contralateral DNLL resulted in a degraded IC FFR. This behaviour suggests that the contralateral DNLL has a principal role in SRM at the level of the IC (Du et al., 2009b, Ping et al., 2008).

3.5.2. Frequency-following response reflects with spatial release from masking at lower SNRs

No significant increase in FFR F0 amplitude was found between co-located and separated conditions at recorded SNRs except for the lowest SNRs when distractors were spatially separated from the target. This finding suggests that the auditory system at the level of the brainstem does not need additional cues to process target stimuli in combination with distractors, provided that the signal is equal to or larger in energy than the distractor stimuli. In other words, the central auditory system processes the target stimulus in the presence of distracter stimuli relied on two strategies. In the first strategy, this system processes the target stimulus based on its overall energy level compared to the overall energy level of distractors as long as the energy level of the target was equal to or higher than the distractors stimuli. In the second strategy, in contrast, when the energy level of the target stimulus was lower than the energy level of distractors stimuli, because the energy level of distractors stimuli buried the information of the target stimulus, the central auditory system changed its signal processing strategy from the energy-level detector to the spatial-cue detector. Moreover, it is physiologically evidenced that the subcortical, pre-attentive processing of voice pitch at the level of brainstem is relatively unaffected by background noise until SNR is decreased to 0 dB or lower (Li and Jeng, 2011). In fact, the steady-state segments of speech, as in vowel- or speech-like stimuli, are relatively resistant to the destructive effects of noise both in animals (Cunningham et al., 2002) and humans (Russo et al., 2004, Song et al., 2011) in positive SNRs. It is assumed, however that, with
decreasing SNR below 0 dB, that the neural mechanisms involved in processing of signal and noise at the brainstem level start to be adversely affected. In behavioural experiments, it has been demonstrated that if there is a spatial separation between noise and signal, the destructive effects of noise are markedly reduced hence the speech reception threshold (SRT) improved by 11-14 dB (Allen et al., 2008). If the neural processing responsible for this effect occurs prior to the generators of the FFR, then the amplitude of the FFR should also be enhanced.

One possible explanation for the observations at low SNRs in our study is via better-ear glimpsing which is based on instantaneous decorrelation, where noise can be substantially weaker in one ear than in the other for relatively short time frames. By attending to the ear with the better SNR in each frequency band and in each fragment in time, the central auditory system combines these portions of the signal to create an optimal representation of the target (Brungart and Iyer, 2012, Glyde et al., 2013a, Zurek, 1993). When SNRs are low, this process potentially has a larger effect as the better-ear glimpsing mechanism may make the difference between being able to track the speech and not track the speech when the SNR is particularly poor in one ear.

3.5.3. Attention modulates FFR F0 amplitude but might not modulate SRM at the level of the brainstem

The results of the current experiment indicated that the FFR F0 amplitude is modulated significantly by selective attention, which is mediated by top-down mechanisms to the target stimuli, which in turn are transferred by bottom-up mechanisms in both spatial conditions. To maintain selective attention, the participants were asked to either relax (or sleep), or focus on the target stimulus and count the number of deviants, which were randomly distributed among the target stimuli. It is acknowledged that by using this paradigm, it is uncertain whether the increase of FFR amplitude can be attributed to an increased state of arousal, an increased attention towards the target stimulus, or a combination of both. This restriction is caused by time limitations with respect to the length of a test appointment. In addition, for clinical purposes, it was reasoned that one is only interested in the question whether the participant (generally a child) can sleep or has to attend to the stimuli during the test.
The current results involving attention indicated that attention to the target stimuli resulted in an enhancement of the FFR F0 amplitude. This finding is in general agreement with other studies (Galbraith and Arroyo, 1993, Galbraith et al., 1998, Galbraith and Doan, 1995, Galbraith et al., 2003). As an example, Galbraith et al. (1998) showed that the F0 response amplitude of vowels /e/ and /a/ becomes significantly larger when subjects attended instead of ignored them. Similar significant effects of attention were found in the current study, although the vowel /u/ used here was different from the stimuli they have utilised in their studies.

During detection, participants selectively attended towards the target stimuli. Selective attention is the mental ability to pick up desired stimuli from a complex acoustic environment by ignoring them (For more information, refer to Fritz et al., 2007). In the context of a discourse-background, accurate perception of the target-foreground is relatively low – especially at low SNRs - and the contribution of bottom-up or stimulus-driven mechanisms that could attract active attention towards the targets is unlikely to be large. Therefore, it is assumed that the top-down mechanisms facilitate the inhibition networks to suppress the distractors and help with concentration on the target streams. Conversely, actively ignoring the target or directing active attention away from the target likely leads to reduced detection probability (Gutschalk et al., 2008). During the active attention trials, actively participating towards the target (in the form of counting a number of deviants) might result in a robust increase of FFR amplitude. A significant effect of attention (as evidenced by Figure 28), regardless of spatial condition or stimulus, was indeed found when participants actively tried to count the number of deviants. Moreover, it is reasonable that spatial separation between target and distractors allowed the participants to engage spatial attention. The spatial attention increased the salience of the target stimulus at the attended location and degraded the distraction caused by the distracted stimuli at the separated location. This conclusion is in accordance with evidence from psychoacoustic experiments (Shinn-Cunningham et al., 2005a).

No interaction effect between SRM and the level of attention was found in the current experiment, as evidenced by Figure 29. That is, the level of attention did not modulate the amount of SRM at the level of the brainstem. Put in other words, top-down attention and bottom-up focusing on a frontal signal operate independently at least at the level of the brainstem. This is promising when evaluating SRM electrophysiologically in
young children. On the other hand, it also might indicate that the frontal preference mechanism that was found here is unrelated to the mechanism that humans generally use to direct their attention to a target. That is, there might be another mechanism that occurs after the site of FFR generation. If so, then the FFR method has no potential for use in detecting spatial processing disorder in children. Unfortunately, no auditory-modality based studies could be found that support either theory. However, this conclusion is in accordance with recent evidence in the visual modality suggesting that between top-down and bottom-up attention there is not any correlation. More specifically, they concluded that there are two attentional mechanisms that operate independently (Pinto et al., 2013). Further, they hypothesised that attention and consciousness are interconnected differently for these two attentional mechanisms. Based on Pinto et al. (2013) and other researchers (Botta et al., 2010, Hein et al., 2006, Yeshurun and Carrasco, 1998), it can be suggested that top-down and bottom-up attentional mechanisms are essentially different processes in the visual system. By extrapolation, and based on the results of the current study, this could be suggested for the auditory modality as well.

3.5.4. Effects of channels

To identify whether different electrode montages might collect different types of information from the possible loci of neural populations responsible for processing of SRM and the possible modulatory effect of top-down attentional mechanisms on this process, two electrode montages have been used: the vertical (channel 1, Fp-C7) and the horizontal (channel 3, Fp-both mastoids) channels. As demonstrated in Figure 32 and Table 1, there is a significant effect between channels 1 and 3 on FFR F0 amplitude across all tested conditions with the larger effect in channel 1. This result is consistent with the previous studies (Galbraith, 1994, Galbraith and Doan, 1995, Galbraith et al., 1998) which showed the largest amplitude in the vertical montage using different stimuli. It is expected that the horizontal montage collects neural activities from more caudal regions of the brainstem and the vertical montage picks up neural activities from more dorsal regions of the brainstem. The horizontal and vertical montages may collect different neural generators. This hypothesis is based on the observed effects on FFR latencies (Galbraith et al., 1998), neural activities from exposed proximal eighth cranial nerve (Møller et al., 1988) and central brainstem regions.
(Scherg and Von Cramon, 1985). Although there is a significant difference between the two channels in terms of FFR amplitude, no statistically significant difference between the two montages were found in terms of SNRs and SRM.

### 3.5.5. Magnitude of effects

To be able to make a direct comparison between the results of this and other studies, changes in FFR amplitude based on dB per nanovolt were calculated. An improvement of 3.3 dB was found in neural processing at the brainstem level when changing from a co-located to a spatially separated condition. The amount of spatial advantage being found electrophysiologically in this experiment is not in agreement with the amount of spatial advantage reported in the psychoacoustic experiments. For instance, using the LiSN-S, Cameron and Dillon (2007a) and Cameron et al. (2011) revealed a spatial advantage of 12.8-12.9 dB. Brown et al. (2010) also reported a 12-dB improvement of SRT as a spatial benefit and Allen et al. (2008) demonstrated a gain of 10.7 to 12.8 dB benefit from spatial separation and finally, Gelfand et al. (1988) showed 5-6 dB in their behavioural counterparts. It is argued that the dissimilarity of target and distractor stimuli, the level of processing in the central auditory nervous system, and the kind of masking (that might be more energetic than informational due to less confusability of target with distractors) are responsible for this discrepancy. However, all these factors should be the subject of further research.

### 3.6. Conclusion

In response to the first main objective of this study, it can be reported that binaural processing relevant to SRM may be reflected in the phase-locked neural activity of a neural population in the brainstem, suggesting that SRM starts early in the central auditory nervous system. SRM is suggested to be most pronounced in relatively challenging listening environments, as the effects were statistically significant at low SNRs only, regardless of whether one actively attended to the target or ignored it. This brings me to the second objective, which investigation has shown that attending to or ignoring from a target stimulus may not play an important role in SRM, at least for frontal signals and at least up to
the region which FFR originates from, which is an important finding when testing children. The estimated benefit from separation (in equivalent input dB SNR) is smaller than commonly measured through behavioural SRM. This suggests that the main effect of SRM might lie higher in the auditory system than the brainstem, or the distractors being used in this study influence the target signals only through energetic masking, or that the remaining part at this level is not measurable electrophysiologically. These questions will hopefully be addressed in the next chapter.
Chapter 4. Investigating spatial processing through the simultaneous recording of the auditory brainstem response (ABR), frequency-following response (FFR) and cortical auditory evoked response (CAEP)
4.1. Preface

The results of the study in Chapter 3 showed that SRM might be reflected in the phase-locked neural activity in the brainstem. SRM was most noticeable at the lowest SNR measured. Further, it could not find evidence that active or passive attention plays a role in SRM at the level of the brainstem. To probe further and higher up the auditory pathway, Chapter 4 will present a second experiment, which addresses the second question of the study. The question is whether electrophysiological measures of neural responses to auditory input identify processing underpinning SRM in either the brainstem or cortex (or both). It also tries to record brainstem and cortical evoked potentials simultaneously. It was assumed that a significant effect of SRM could be measured in both the brainstem (through the ABR and FFR), and the auditory cortex (through the CAEP).

4.2. Introduction

As already shown in the literature review of Chapter 2, and the first study in Chapter 3, which revealed that FFRs could reflect SRM at the level of the brainstem electrophysiology could be a useful technique to investigate how the central auditory system processes a target signal that is spatially mixed with different distractors, and how the processing of the target signal is improved under low SNRs. To probe further, the electrophysiological tools, which will be used in this second experiment, are the auditory brainstem response (ABR), the frequency-following response (FFR), and the cortical auditory evoked potential (CAEP).

To summarise the electrophysiological in-depth overview of Chapter 2, the ABR is an electrical activity from the brainstem evoked by a brief stimulus. It is an early response and occurs within 10-15 ms after stimulus onset and encompasses several waves which are labelled by a roman numeral system from I to VII (Guerreiro and Ehrenberg, 1982). It is generated in different parts of the brainstem. The neural generators of wave I, II, III, IV, and V have been proposed to originate from the distal part of the VIII cranial nerve (Luders et al., 1988, Grandori, 1986, Ino and Mizoi, 1980), the proximal part of the VIII cranial nerve (Møller et al., 1988, Møller and Jannetta, 1982, Møller et al., 1995), the caudal pons (Achor
and Starr, 1980, Legatt et al., 1986, Grandori, 1986, Scherg and Von Cramon, 1985), the superior olivary complex (SOC) (Møller et al., 1995, Moore, 1987) the inferior colliculus (IC) (Lev and Sohmer, 1972, Scherg and Von Cramon, 1985), and the rostral part of lateral lemniscus (LL) or/and the IC (Hashimoto et al., 1981, Møller et al., 1995), respectively.

The FFR is a sustained response of neural activities to low-frequency stimuli. It is phase-locked to the individual cycles of a stimulus waveform or its periodic envelope (Gardi et al., 1979b, Moushegian et al., 1973). The neural generators of the FFR has been proposed to be in the rostral part of brainstem (Smith et al., 1975).

The CAEP is a late response which occurs 50-200 ms after stimulus onset and is evoked to onset, offset, and changes of a stimulus (Lightfoot, 2010). It has different components in adults which are known as P1, N1, P2, and N2 and is generated in different parts of the auditory cortex (Stapells, 2002). The generators of P1 have been proposed to be in the primary auditory cortex particularly in the Heschl’s gyrus (Huotilainen et al., 1998), the secondary auditory cortex (Ponton et al., 2000) and other areas include the planum temporal, hippocampus, and lateral temporal (Howard et al., 2000, Liegeois-Chauvel et al., 1994). The N1 component is thought to have several neural generators in both primary and secondary auditory cortices (Howard et al., 2000, Näätänen and Picton, 1987, Scherg et al., 1989, Stillman et al., 1978, Vaughan and Ritter, 1970). The P2, likewise, has several neural generators which are located in the primary auditory cortex (Baumann et al., 1990) and the secondary auditory cortex (Hari et al., 1987).

Several studies have attempted to investigate whether the activities of neural population underling ITD and ILD cues can be represented via ABR components (Brantberg et al., 1999, Furst et al., 1985, Gerull and Mrowinski, 1984, Jones and Van der Poel, 1990, McPherson and Starr, 1995, Riedel and Kollmeier, 2002). For more detail about these studies, readers are referred to Chapter 2. These studies showed that the effects of changing ITD and ILD cues considering lateralisation paradigm indeed could be measured at the level of the brainstem using the ABR components as a potential objective measure of theses spatial cues. As an example, Furst et al. (1985) and McPherson and Starr (1995) reported that ABR components (in the form of the binaural interaction component) systematically changed as a function of ILD and ITD. Detecting changes in these spatial cues via ABR showed that this objective measure would be able to provide useful information.
regarding binaural processing, and particularly regarding spatial processing. More specifically, as the generators of the ABR wave V are thought to be anatomically situated in the LL and/or the IC, one could expect to collect information regarding spatial processing up to the ABR wave V generators. Hence, it is plausible in the current study to investigate SRM using the time-locked wave V of the ABR.

Similarly, binaural processing has been investigated by the FFR utilising spatial cues, ITD and ILD in lateralisation paradigms (Ballachanda and Moushegian, 2000, Clark et al., 1997, Krishnan and McDaniel, 1998). Their encouraging results, as outlined in Chapter 2, supported the use of the FFR in Chapter 3 and the current chapter to investigate SRM.

The main body of research on the cortical response involving spatial processing signal and noise has focused so far on three aspects. First, the effect of lateralisation and localisation utilising individual ILD and ITD cues on CAEPs has been investigated (Butler, 1972, Edmonds and Krumbholz, 2014, Picton et al., 1991, Wambacq et al., 2007). Second, a few experiments have focused on the effects of distractors on the target signal using the cortical response (Lewald and Getzmann, 2015, Zhang et al., 2014). Third, other research has investigated the representation of neural populations in the auditory system while processing signal and noise. The aim was to explain how the auditory system in humans is able to extract the voice-pitch of speech information in the presence of background noise even when SNR is as low as -10 dB (Kong and Zeng, 2006), and hence to improve our understanding of the underlying mechanisms involved in speech perception in noisy environments. Generally, noise maskers increase the latency and decrease the amplitude both in subcortical and cortical level (Beattie and Kennedy, 1992, Beattie et al., 1994, Burkard and Hecox, 1983, Li and Jeng, 2011, Martin et al., 1996, Martin and Stapells, 2005).

Although the separate electrophysiological study of the behaviour of the auditory system to the three individual aspects mentioned above sheds some light on the underlying auditory mechanisms, taking into account the combined effects of spatial location of noise and signal at different SNRs can help to understand the auditory mechanisms underlying SRM.

This second study asks one main research question that currently remains unresolved. It aims to answer

- Whether electrophysiological measures of neural responses to auditory input can identify processing underpinning SRM in either the brainstem or cortex
(or both). More specifically, can the significant results at -5 dB SNR of Chapter 3 be reproduced at the level of the brainstem (for the FFR), and does SRM result in an increase of ABR and CAEP amplitudes and a decrease of ABR and CAEP latencies?

The current experiment investigates whether SRM can be recorded electrophysiologically in normal-hearing adults during spatial processing of target tone-complexes spatially co-located with, and separated from, tone-complex distractors employing FFRs, ABRs, and CAEPs. Using tone-complex stimuli, any possible confounding factors are avoided that are induced by the use of language stimuli such as semantic effects and speech understanding.

4.3. Materials and methods

4.3.1. Participants

Thirteen adults participated in this study. The participants were 7 females and 6 males ranging in age 20-44 years (mean age: 31.69, SD: 6.5 years). The participants were selected from a list of volunteers at the National Acoustic Laboratories (NAL) and from students from Macquarie University via public announcement, hence non-probability sampling. They had normal hearing bilaterally from 125-8000 Hz (with hearing thresholds ≤20 dB HL). None of the participants had a history of otologic and neurologic disorders. Participants received a written informed-consent form. For this study, the Australian Hearing Human Research Ethics Committee (AHHREC) approved the procedures and protocols. All subjects were reimbursed for their travel expenses.

4.3.2. Stimuli

The stimuli in this study closely followed the design of the Listening in Spatialised Noise – Tonal (LiSN-T) test. The LiSN-T aims to evaluate spatial auditory processing in both normally hearing individuals and those with spatial processing disorder (SPD) psychoacoustically (Buchholz et al., 2013). The development of the LiSN-T was inspired by the Listening in Spatialised Noise - Speech (LiSN-S) test but exclusively relies on non-
language stimuli. LiSN-S is an adaptive virtual test that has been developed by the National Acoustic Laboratories (NAL) for evaluating the understanding of speech in presence of spatialised babble noise (Cameron and Dillon, 2008).

Stimuli in the current experiment, both targets and distractors, were presented in J sequential blocks (Figure 33). Each block $j$ ($j = 1, \ldots, J$) contained the same, identical pattern of 11 target stimuli, with a random selection of 11 distractor stimuli. The target stimuli (identical for each block) were composed of a series of 11 tone complexes (TCs). The individual TCs had a duration of 30 ms with 5 ms long raised cosine-shaped onset and offset ramps, and with a 30 ms interstimulus interval (ISI). The $i^{th}$ TC ($i = 1, \ldots, 11$) started at $15+(i-1)*60$ ms and ended at $45+(i-1)*60$ ms from the onset of each block. These TCs were assumed to evoke the ABRs and FFRs. The blocks hence were $11\times(15+30+15) = 660$ ms long, were repeated every 1200 ms, and the blocks’ onsets were supposed to evoke the CAEPs.

Two block patterns were designed: the flat and the downward staircase frequency pattern Figure 33. Based on the pilot tests by Buchholz et al. conducted at the National Acoustic Laboratories (personal communication), it was concluded that a downward staircase frequency pattern created a significantly larger spatial advantage than a flat frequency pattern in the separate blocks. To compare their effects, both patterns were included. The flat blocks had TCs with a constant fundamental frequency $F_0$ of 325 Hz, and its harmonics up to 6 kHz. The level of the harmonic tones reduced by 6 dB per octave to resemble the average speech spectrum (Figure 34 and Figure 35). The fundamental frequencies of the TCs in the staircase blocks went downward from 475 to 175 Hz in steps of 30 Hz, again with harmonics up to 6 kHz (Figure 33 and Figure 35).

The distractors were blocks of TCs similar to the structure of the target stimuli, but randomised in their $F_0$ distribution from 100 to 550 Hz, symmetrically around the target $F_0$ range (Figure 33). This increase in the frequency range was adopted to minimise the cue that otherwise would have been provided by the target signal always ending at the highest frequency. The distractors’ onset times were jittered between -15 and 15 ms relative to the onset of the target TC. Ongoing distractors were used to minimise the likelihood of the distractors giving rise to cortical responses.
Figure 33. Eleven 30-ms long target tone complexes (green colour) are presented together with jittered 30-ms long distractors (red colour). These target-distractor combinations are then followed by nine distractors. The flat targets have a constant F0 of 325 Hz (top). The downward staircase targets have F0s that drop from 475 to 175 Hz in steps of 30 Hz (bottom). The F0s of the distractors are randomised between 100 and 550 Hz.

There initially was a difference between the behavioural and electrophysiological stimulus paradigms in the way that the distractor stimuli were continuous in the
electrophysiological test (to avoid CAEPs evoked by the onsets of the distractors) but not in the behavioural part (the LiSN-T). In the behavioural part, the distractors started simultaneously with the target stimuli (i.e., simultaneous masking). For this reason, nine F0-randomized distractors were added (i.e., fringed masking) before and after the actual distractors (those randomised with the target stimuli) in the behavioural version designed for this experiment. The original LiSN-T (without fringed distractors) was compared psychoacoustically with the fringed version in a behavioural pilot study. In this pilot study, four participants were recruited (age range: 20-34, mean 29.5, SD 6.3 years) with 1 female and three males. They were tested under headphones. Subjects were presented a 3 alternative forced choice paradigm, with one presented option containing targets and distractors, and the two remaining options only containing distractors. They had to indicate the option where they could hear the target. SNRs were automatically adapted depending on the answer of the subject, and a detection threshold for co-located and separated spatial conditions was determined for both flat and staircase stimulus paradigms. Stimuli were identical to the stimuli outlined in this experiment. The remainder of the protocol and subsequent calibration was identical to the relevant LiSN-T study (Buchholz et al., 2013). The means and SDs of the detection thresholds for both co-located and separated conditions in flat frequency were -11.2±4.0 dB and -21.8±2.0 dB and were -5.4±3.5 dB and -18.6±2.9 dB in staircase frequency respectively. No significant differences in SRM were found comparing both non-fringed and fringed versions in flat ($F_{(1, 3)}=0.64$, $p=0.48$) and staircase ($F_{(1, 3)}=2.65$, $p=0.20$) stimulus paradigms.

The target and distractor stimuli, as shown in Figure 34 and Figure 35 were convolved with non-individualized head-related transfer functions (HRTFs) measured on a Knowles Electronics Manikin for Acoustic Research (KEMAR) manikin. The targets were located at 0⁰ azimuth for all recordings. The distractor stimuli were located in a co-located (at 0⁰ azimuth) and spatially separated (at ±90⁰ azimuth) condition. This stimulus manipulation induced the perception that the target and distractor TCs in the co-located condition were at the midline location. In the separated condition, the perception of targets remained in the midline, but distractors moved away from the midline to the sides. Stimulus presentation was controlled by MATLAB software (Math Works Inc, 2012) through a high-fidelity multichannel RME soundcard (ADI-8 QS) and electromagnetically shielded Etymotic
Figure 34. The waveform, the spectrogram and spectra of TC target stimulus (top) in flat frequency paradigm. The waveforms, spectrograms and spectra of TC discourses in co-located left and right channel (middle panels), separated left and right channel (bottom panels).
Figure 35. The waveform, the spectrogram and spectra of TC target stimulus (top) in staircase frequency paradigm. The waveforms, spectrograms and spectra of TC discourses in co-located left and right channel (middle panels), separated left and right channel (bottom panels).
Research ER-3A insert earphones with 28 cm air tubes. Stimuli were calibrated at 60 dB HL using a IEC126 HA2 2-cc coupler, incorporating a 1-inch 4144 microphone, a 1-to-1/2 inch DB0375 adaptor, and a 30 second averaging time window on a Brüel and Kjær 2636. The target stimuli were presented at 60 dB SPL bilaterally, with SNRs of -5, 0, 5, 10 and 15 dB.

The different SNRs were chosen to simulate routine listening environments by altering the degrees of difficulty in understanding targets in the presence of distractors. This wide range of SNRs included low SNRs to examine the resistance of the brainstem against relatively loud distractors, and to find significant effects of SRM at low SNRs similar to the first experiment of this thesis, which is reported in Chapter 3. High SNRs also considered to avoid CAEPs to be obliterated because of poor SNR conditions, as CAEPs are relatively sensitive to noise levels (Billings et al., 2009). When distractors were presented simultaneously to both ears, the SNR dropped on average by 3 dB.

4.3.3. Procedure

The experiment was split into two appointments. In one session the flat frequency stimulus and in the other session the staircase frequency stimulus was tested. The order of conditions and appointments were randomized and counterbalanced across participants to control for order effects. The experiment divided into two portions, the behavioural, and the electrophysiological.

4.3.3.1. Behavioural evaluation

Otoscopy and pure tone audiometry were performed prior to the behavioural test. The behavioural test included the Listening in Spatialised Noise-Tone (LiSN-T) test (Buchholz et al., 2013) with both flat and downward staircase paradigms. For the behavioural assessment of the LiSN-T, the same flat and staircase stimuli were used as in the electrophysiology test, as described in Section 4.3.2. A three-alternative forced-choice pattern discrimination task was presented to the participants, with one of the three options containing the flat or staircase target. The stimuli were presented under headphones with the computer providing instructions to the participants. The participants were informed that three boxes would appear on the computer screen. These boxes turned sequentially
accompanied by a stimulus pattern. The subject was then asked to identify the box that contained the target stimuli. The level of the target sound decreased or increased depending on whether the subject correctly guessed the box containing the target or not. When the participants hit the correct box containing the target, the computer decreased the SNR with a 2 dB-step interval of target stimulus level. This adaptive procedure determined a detection threshold as the mean of 3 reversals.

4.3.3.2. **Electrophysiological evaluation**

Participants were seated in a comfortable recliner in a sound-proofed electromagnetically-shielded and dimly lit booth. They were instructed to be relaxed as much as possible during recording and to ignore the acoustic stimuli while watching a silent closed-captioned movie. To avoid any electromagnetic interference, all unused equipment and power outlets were turned off.

Three gold-plated electrodes were placed onto the participant’s head following preparation of the skin by rubbing with cotton tissue and Nuprep gel. The active electrode was located on the forehead on the hairline (Bell et al., 2004) and the reference electrode was located on the right mastoid (Klem et al., 1999). The common electrode was placed on the lower forehead (Fpz). The correct locations were verified using a tape measure. To ensure good contact between electrodes and skin, all electrodes required a reasonable amount of electrode paste. Impedances between active and ground electrodes, and reference and ground electrodes were kept ≤ 1 kΩ. Impedance of electrodes was checked before and after the recordings, and during pauses between the recordings. Electrodes were reapplied if necessary.

Electroencephalography (EEG) signals were obtained using an Intelligent Hearing Systems (IHS) measurement system. The recorded channel was analogue band-pass filtered into two channels to obtain cortical and brainstem recordings simultaneously. The cortical channel was filtered between 1 and 30 Hz (6 dB/octave roll-off) and amplified with a gain of 5,000. The brainstem channel was filtered between 30 and 3000 Hz (6 dB/octave roll-off), with a 50 Hz notch filter, and had an amplification of 150,000. The analogue-to-digital sampling rate was 10 kHz.
In total, when considering two frequency paradigms (flat and staircase), two spatial conditions (co-location and separation), five SNRs (-5, 0, 5, 10, and 15 dB) and test re-test conditions, 40 stimulus conditions were evaluated for each participant. Per condition, 200 blocks comprising 11 TCs each were presented with 1200 ms between block onsets. The approximate time to complete the whole sessions took 2 sessions of 2 hours each for a total of 4 hours. The data analysis was conducted off-line.

### 4.3.4. Data analysis

The top 10 percent of epochs with the largest absolute maximum values were excluded off-line for each recording condition. For the analysis of brainstem responses (ABR and FFR), 1,980 accepted epochs were combined using noise-power weighted averaging. The FFR could only be derived for the flat stimuli, as the number of TC presentations (180) for each F0 in the staircase paradigm was not sufficient to derive a clear FFR. For the flat stimuli, a Fast-Fourier Transform (FFT) was applied to the window from 7.7 to 37.5 ms post stimulus onset for the processing of the FFR in the frequency domain with a bin width of 25 Hz. This bin width seemed optimal as it is about the inverse of the stimulus duration, hence it would capture most of the energy of each FFR component. For the FFR, the amplitude at F0 = 325 Hz was determined (Figure 36).

In the case of the staircase stimuli, different F0s of the TC stimuli (475 to 175 Hz) resulted in different stimulus onset times and latencies of the first maximum or minimum peaks, each evoking a different wave V of the ABR. As a result, the evoked wave V of the ABR had different latencies, causing the ABRs of the different TCs to superimpose and destructively interfere with each other. To overcome this limitation and to create a robust wave V of the ABR representing all TCs equally, the delays of the stimulus onsets of each TC were determined and compensated for by adding or subtracting a specific delay. More specifically, all 11 ABR waveforms were lined up and combined using the onset times of the 11 separate staircase TCs to establish a single compensated waveform. First, the delay in milliseconds between the start of the wave file and the first maximum or minimum peak of each tone complex was determined. This delay incorporated the delay of the ER-3A insert tubes (1.8 ms corresponding to 28 cm) and some additional digital-to-analogue conversion
delay. Given the TCs had different F0s and their harmonics, the first maximum (or minimum) peak occurred at different delays. The delays of TCs in the downward staircase paradigm (475, 445, 415, 385, 355, 325, 295, 265, 235, 205, 175 Hz) were 2.2, 2.3, 2.5, 2.6, 2.9, 3.2, 3.5, 3.9, 4.4, and 5 ms respectively. Second, all stimulus onsets were adjusted relative to the stimulus onset associated with the centre TC of the staircase, i.e. with F0 = 325 Hz, corresponding to a delay of 3.2 ms. This resulted in adjustments of -1.0, -0.9, -0.6, -0.5, -0.3, 0.0, 0.3, 0.7, 1.2 and 1.8 ms respectively. Finally, the corresponding ABRs that were evoked by the 11 TCs were compensated by the delayed time adjustments calculated above. The flat stimuli were delay compensated as well, but only in function of the onset time of the TC with an F0 of 325 Hz, or 3.2 ms. This resulted in an identical delay time compensation for all flat TC stimuli, making sure they had the exact same latencies as the staircase TCs, allowing comparison between the two paradigms.

ABR wave V was determined manually on the test-retest grand average in the range of 8 to 13 ms (a range determined using the grand average in Figure 36), using extra information in the form of test and retest traces to indicate repeatability, and the available waveforms at higher SNRs. In addition, for the flat stimuli, wave V was determined as the first positive peak of any identifiable FFR (Figure 36). The examiner was blinded from spatial condition during wave V identification. Wave V latency (in ms) was determined as the time from stimulus onset. Wave V amplitude (in µV) was determined as the difference between wave V and the first negative wave following wave V, commonly called the SN10.

For the analysis of the CAEP results, the number of accepted epochs was equal to 180 across all conditions. The recorded epoch time consisted of a 200 ms prestimulus baseline followed by 1500 ms post stimulus onset. A positive peak (P1) and a negative peak (N1) were identified. It has to be noted however that these CAEP waveforms were not akin to the commonly known positive and negative onset deflections (a P1-N1-P2 pattern) in response to one single stimulus. More specifically, the reported CAEPs in this experiment were generated by a series of consecutive TCs, with each TC individually evoking their own CAEP, resulting in a combined CAEP consisting of multiple overlapping responses. Latency
Figure 36. Grand average (thick line) and individual traces (thin lines) over all 13 subjects of the 15 dB SNR co-located conditions. The top left corner shows the averaged brainstem response to the flat stimulus condition, with an indication of the ABR latency and the FFR. The top right corner shows its FFT, with an indication of the FFR amplitude. The bottom left corner shows the averaged brainstem response to the staircase stimulus condition. The FFRs to the different tone complexes are averaged out. The bottom right corner shows the CAEP to the flat stimulus condition, with indications of a positive wave at 80 ms latency and a negative wave at 180 ms latency.

Latencies of the CAEP components were determined relative to signal onset and the wave amplitudes were specified relative to baseline. The amplitudes and latencies of the CAEPs peaks were obtained for both flat and staircase frequency paradigms. Peak amplitudes and latencies were determined by automatic peak picking. The automatic peak picking was performed because of poor CAEP SNRs, which made manual peak picking highly challenging. Latency ranges used by the automatic peak picking were determined beforehand using the grand averages over all subjects, shown in Figure 37. CAEP peak picking occurred according to the following rules:

- P1 maximum had to occur between 50 and 300 ms
- N1 minimum had to occur between 150 and 500 ms
- N1 always had to follow P1
- N1 always needed to follow within 200 ms after P1. The reasoning behind this is that the main frequency content of a CAEP is around 5 Hz, such that a full cycle is 200 ms. This way a negative peak generally was avoided to be peak picked after two positive peaks.
Due to high variability of the P2 waveform, it was not feasible to reliably identify this peak in the current experiment.

### 4.3.5. Statistical analysis

It will become clear during the presentation and the analysis of the recorded waveforms in Sections 4.4.2 and 4.4.4 that these waveforms are particularly noisy, especially for the ABR and CAEP components. This likely is caused by the specific nature of the stimuli and the relatively low acoustic SNRs the waveforms were recorded at. To improve the quality of the analysed waveforms, test and retest data were averaged prior to statistical analysis. This averaging process was considered acceptable, as no statistically significant effect of (re)test was found for any of the measures analysed below (for statistical values, please refer to Section 4.4.2 and 4.4.4).

Statistical analyses were conducted with STATISTICA software package (version 12 SP3). After test-retest averaging, for the analysis of the FFR F0 amplitude, the amplitude and latency of the ABR and cortical P1 and N1 components, a three-way (2x2x5) analysis of variance (ANOVA) was performed with target frequency paradigm (flat versus staircase), spatial condition (co-located versus separated), and target SNR (-5, 0, 5, 10 and 15 dB) as repeated-measures variables. The a priori level of significance was chosen to be $p=0.05$. When a post hoc power analysis was required, a Tukey’s honest significant difference (HSD) tests were performed.
Figure 37. Grand averages from 13 subjects for -5, 0, 5, 10 and 15 dB SNR for flat stimulus co-located (top left), flat stimulus separated (top right), staircase stimulus co-located (bottom left) and staircase stimulus separated (bottom right).

### 4.4. Results

This section evaluates the behavioural and the electrophysiological effects of spatial separation of the targets from the distractors. In the behavioural correlate (Section 4.4.1), the test-retest results are presented for the Listening in Spatialised Noise-Tone (LiSN-T) test, which was conducted prior to electrophysiological testing. In the electrophysiological correlate (Sections 4.4.2 to 4.4.4), the effect was investigated of frequency paradigm of the target (flat and downward staircase), spatial condition (co-located and separated distractors from target), and target SNR (-5, 0, 5, 10, and 15 dB) on the amplitude and latency of the ABR, the amplitude of the FFR, and the amplitude and latency of the CAEPs in response to the target stimuli, after taking the waveform grand average of the test-retest data.
4.4.1. Behavioural results (LiSN-T)

The reception thresholds in the co-located condition were -5.0 dB (SD 7.2 dB, range -21.8 dB to 3.3 dB). For the separated condition, these thresholds were -19.1 dB (SD 3.8 dB, range -24.3 to -9.0 dB). The spatial advantages, which were given by the difference between the behavioural thresholds measured in spatial co-located and separated conditions, were on average 10.7 dB (SD 7.8 dB) and 14.0 dB (SD 6.0 dB) for flat and staircase stimuli, respectively. For both flat and staircase paradigms, a two-way analysis of variance (ANOVA) was performed on the spatial condition (co-located versus separated) and test condition (test versus re-test) as repeated-measures variables. A significant effect of SRM was found on the measured thresholds both for flat ($F_{(1, 12)} = 24.64, p < 0.001$) and staircase stimuli ($F_{(1, 12)} = 71.44, p < 0.001$). These results show that for both stimulus paradigms, spatial separation of the target from its distractors indeed improves behavioural reception thresholds.

4.4.2. Auditory brainstem response

For all ABR measures (wave V amplitude and latency), no significant main effect between test and retest was found, allowing test-retest averages of the waveforms (Table 2).

4.4.2.1. Wave V amplitude

Figure 38, Figure 39, Figure 40 and Figure 41 show wave V amplitudes versus SNR for 9 spatially co-located and separated conditions, with a flat (Figure 38 and Figure 39) and downward staircase (Figure 40 and Figure 41) target frequency paradigm respectively.
Chapter 4: Investigating spatial processing through ...

Figure 38. Grand averages waveforms of FFR F0 amplitude overlaid in 13 participants in time domain at 5 tested SNRs (from left to right side, are -5, 0, 5, 10, and 15 dB) in co-located and separated spatial conditions for flat frequency paradigm. ABR wave V is considered as the first positive of the first FFR positive waveform.

Figure 39. Means and standard errors of ABR wave V amplitude in spatially co-located and separated conditions as a function of target SNR, averaged across test and re-test when using a flat target frequency paradigm.
Chapter 4: Investigating spatial processing through ...

Figure 40. Grand averages waveforms of FFR F0 amplitude overlayed across participants in time domain at 5 tested SNRs (from left to right side, are -5, 0, 5, 10, and 15 dB) in co-located and separated spatial conditions for staircase frequency paradigm. ABR wave V is considered as the first positive of the first FFR positive waveform.

<table>
<thead>
<tr>
<th>SNR (dB)</th>
<th>Co-located</th>
<th>Separated</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.06</td>
<td>0.08</td>
<td>0.10</td>
</tr>
<tr>
<td>0.12</td>
<td>0.14</td>
<td>0.16</td>
</tr>
<tr>
<td>0.18</td>
<td></td>
<td>0.20</td>
</tr>
</tbody>
</table>

Figure 41. Means and standard errors of ABR wave V amplitude in spatially co-located and separated conditions as a function of target SNR, averaged across test and re-test when using a staircase target frequency paradigm.

4.4.2.1.1. Main effect of SRM on wave V amplitude

Based on the data shown in Figure 39 and Figure 41, no significant main effect of SRM was found on the amplitude of ABR wave V (Table 2). These results suggest that spatial separation cannot be recorded electrophysiologically using the ABR wave V amplitude.
4.4.2.1.2. Main effect of SNR on wave V amplitude

A significant main effect of stimulus SNR on wave V amplitude was found (Table 2). With reducing level of the background distractors, the amplitude of wave V increased accordingly.

4.4.2.1.3. Main effect of target frequency paradigm on wave V amplitude

A significant difference was found between flat and staircase frequency paradigms (Table 2) resulting in a larger wave V amplitude evoked by the flat paradigm when compared to the staircase paradigm.

4.4.2.1.4. Interaction effect between SRM and SNR on wave V amplitude

No significant interaction effects of SRM and SNR were found on the amplitude of the ABR wave V for flat frequency paradigm however for staircase frequency paradigm this interaction effect was significant (Table 2).

4.4.2.1.5. Interaction effect between SRM and target frequency paradigm on wave V amplitude

The interaction effects between SRM and target frequency paradigm on the ABR wave V were not significant suggesting that flat and staircase frequency paradigm could not have different effect on wave V amplitude in spatial conditions (Table 2).
Table 2. Results of repeated measures analysis of amplitude and latency of the wave V ABR, FFR F0 amplitude, and amplitudes and latencies of the CAEP components P1 and N1 for SNR, SRM and interaction between SRM and SNR using flat and staircase frequency paradigms (Stim). Where no significant interaction between paradigm, SRM and SNR was found, statistics are reported with the paradigms combined.

<table>
<thead>
<tr>
<th>Test (3-way)</th>
<th>SRM × SNR</th>
<th>SRM × Stim</th>
<th>SNR × Stim</th>
<th>SRM × SNR × Stim</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABR (wave V)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amplitude (µV)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F(1,12) = 0.36, p = 0.56</td>
<td>F(4,48) = 12.3, p &lt; 0.0001</td>
<td>F(1,12) = 48.8, p &lt; 0.0001</td>
<td>F(4,48) = 1.65, p = 0.18</td>
<td>F(1,12) = 0.07, p = 0.80</td>
</tr>
<tr>
<td>Latency (ms)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F(1,12) = 19.6, p = 0.0008</td>
<td>F(4,48) = 26.8, p &lt; 0.0001</td>
<td>F(1,12) = 4.02, p = 0.068</td>
<td>F(4,48) = 2.35, p = 0.067</td>
<td>F(1,12) = 1.24, p = 0.29</td>
</tr>
<tr>
<td>FFR Amplitude (µV)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F(1,12) = 0.05, p = 0.82</td>
<td>F(4,48) = 9.80, p &lt; 0.0001</td>
<td>-</td>
<td>F(4,48) = 2.53, p = 0.050</td>
<td>-</td>
</tr>
<tr>
<td>CAEP Amplitude (µV)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F(1,12) = 0.49, p = 0.50</td>
<td>F(4,48) = 7.37, p = 0.0001</td>
<td>F(1,12) = 4.80, p = 0.049</td>
<td>F(4,48) = 0.71, p = 0.59</td>
<td>F(1,12) = 8.83, p = 0.012</td>
</tr>
<tr>
<td>N1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F(1,12) = 0.93, p = 0.35</td>
<td>F(4,48) = 0.80, p = 0.53</td>
<td>F(1,12) = 13.4, p = 0.003</td>
<td>F(4,48) = 0.77, p = 0.55</td>
<td>F(1,12) = 5.63, p = 0.035</td>
</tr>
<tr>
<td>CEP Latency (ms)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F(1,12) = 0.68, p = 0.43</td>
<td>F(4,48) = 1.81, p = 0.024, p = 0.88</td>
<td>F(1,12) = 2.02, p = 0.11</td>
<td>F(4,48) = 0.23, p = 0.64</td>
<td>F(4,48) = 1.04, p = 0.40</td>
</tr>
<tr>
<td>N1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F(1,12) = 0.00, p = 0.96</td>
<td>F(4,48) = 2.88, p = 0.032</td>
<td>F(1,12) = 0.16, p = 0.70</td>
<td>F(4,48) = 1.24, p = 0.31</td>
<td>F(1,12) = 0.02, p = 0.90</td>
</tr>
</tbody>
</table>

4.4.2.2. Wave V latency

Figure 42 and Figure 43 show ABR latencies versus SNR for spatially co-located and separated conditions, with a flat (Figure 42) and downward staircase (Figure 43) target frequency paradigm respectively.
Figure 42. Means and standard errors of ABR latency in spatially co-located and separated conditions as a function of target SNR, averaged across test and re-test when using a flat target frequency paradigm.

Figure 43. Means and standard errors of ABR latency in spatially co-located and separated conditions as a function of target SNR, averaged across test and re-test when using a staircase target frequency paradigm.

4.4.2.2.1. Main effect of SRM on wave V latency

A significant main effect of SRM was found on the latency of the ABR wave V with the latency decreasing when spatially separating distractors from the target (Table 2). A mean latency difference of 0.62 ms (SE = 0.14 ms) was found between co-located and separated conditions with a 95% confidence interval of 0.31 to 0.92 ms. These results
suggest that spatial separation can be recorded electrophysiologically using the ABR wave V latency.

4.4.2.2.2. Main effect of SNR on wave V latency

A significant main effect of stimulus SNR on wave V latency was found (Table 2). With reducing level of the background distractors, the latency of the ABR increased accordingly. The means and standard errors of wave V latency for the co-located condition at -5 and 15 dB SNR were 11.10±0.60 and 10.21±0.61 ms respectively, and 10.57±0.67 and 9.62±0.37 respectively for the separated condition. This corresponded to an average decrease of 0.045 ms per dB increase of target SNR in the co-located condition (by averaging 0.89 ms over 20 dB).

The co-located condition can be used as a reference condition to calculate the absolute benefit of spatial separation with wave V latency as an objective measure. A mean latency difference of 0.62 ms was found when spatially separating the distractors from the target. Combined with the numerical information above (0.045 ms per dB), this results in an absolute benefit of 13.8 dB (95% confidence interval of 6.9 and 20.4 dB) of SRM when using wave V latency as a measure in terms of the required increase of stimulus SNR to obtain the same effect of SRM.

4.4.2.2.3. Main effect of target frequency paradigm on wave V latency

No significant difference was found between flat and staircase frequency paradigms in ABR wave V latency (Table 2), suggesting that the delay compensation procedure outlined in Section 4.3.5 performs as expected.

4.4.2.2.4. Interaction effect between SRM and SNR on wave V latency

No significant interaction effects were found on the latency of the ABR wave V while using flat frequency paradigm but there is a tendency towards significant when using staircase frequency paradigm (Table 2). These results suggest that flat and staircase frequency paradigms might not create same spatial-cue strengths while reflecting via wave
Chapter 4: Investigating spatial processing through ...

V latency. Moreover, latency of the ABR wave V may sensitive to reflect SRM compared to amplitude.

4.4.2.2.5. Interaction effect between SRM and target frequency paradigm on wave V latency

No significant interaction effects between SRM and target frequency paradigm were found on the latency of the ABR wave V suggesting that flat and staircase frequency paradigm could not have different effect on wave V latency in spatial conditions (Table 2).

Overall, these results suggest that the ABR waveform latency might have a potential to be an objective tool for reflecting SRM if appropriate stimuli will be used for eliciting ABR waveform.

4.4.3. Frequency-following response (FFR)

Reporting of the FFR results is limited to the flat frequency paradigm. Because of the 11 tone complexes with 11 different fundamental frequencies in the staircase frequency paradigm, obtaining a clear grand average of the FFRs was not possible with the limited number of 180 accepted presentations at each fundamental frequency. In addition, FFR results will only be reported through the FFR amplitude. As shown in Figure 36, the ABR latency is defined as the delay of the first positive peak of the FFR. Figure 45 shows the FFR amplitude versus SNR for spatially co-located and separated conditions for a flat frequency paradigm only. No significant difference was found between test and re-test (Table 2).

4.4.3.1. Main effect of SRM on F0 amplitude

No significant effect of SRM on FFR F0 amplitude was found while grouping all data across all SNRs, test, and retest (Table 2). This result suggests that if different SNRs would have different potential to show SRM, pooling the data of different SNRs together offsets each other effects hence no SRM could be detected via FFR F0 amplitude.
4.4.3.2. **Main effect of SNR on F0 amplitude**

Figure 44 and Figure 45 show that with increasing target stimulus SNR, the amplitude of the fundamental frequency of the FFR increased significantly (Table 2). For the co-located condition, the means and standard errors of FFR amplitude at -5 and 15 dB SNR were

![Grand averages waveforms of FFR F0 amplitude overlaid participants in frequency domain at 5 tested SNRs](image1)

**Figure 44.** Grand averages waveforms of FFR F0 amplitude overlaid participants in frequency domain at 5 tested SNRs (from left to right side, are -5, 0, 5, 10, and 15 dB) in co-located and separated spatial conditions for flat frequency paradigm.

![Means and standard errors of FFR F0 amplitude](image2)

**Figure 45.** Means and standard errors of FFR F0 amplitude in spatially co-located and separated conditions as a function of stimulus SNR when using a flat target frequency paradigm.
18.78±2.10 and 45.59±6.47 nV respectively, and for the separated condition were 24.59±2.50 and 40.11±5.17 nV respectively. These results corresponded to an average increase (26.81 nV over 20 dB) of 1.34 nV per dB increase of the target SNR in the co-located condition.

### 4.4.3.3. Interaction effect between SRM and SNR on F0 amplitude

A significant interaction was found between spatial condition and stimulus SNR (Table 2). This warranted a further breakdown of the spatial separation at separate SNRs. Post hoc comparison employing the Tukey’s HSD test showed that the mean difference between separated and co-located conditions was significant (p = 0.0088) at -5 dB SNR with an amplitude difference of 5.81 nV, and a 95% confidence interval of 1.75 to 9.86 nV. The mean difference was close-to-significant (p = 0.054) at 0 dB SNR with an amplitude of 7.39 nV, and a 95% confidence interval of -0.14 to 14.91 nV. There were no significant differences between co-located and separated conditions at other SNRs.

In the -5 dB SNR stimulus condition, the effect of SRM corresponds to an absolute increase of 4.34 dB (5.81 nV SRM/ 1.34 nV per dB) in terms of the required increase of stimulus SNR to obtain the same effect of SRM using F0 amplitude as a measure.

Overall, these results suggest that SRM is represented at the level of the brainstem where the generators of FFR are situated, but only at the lowest SNR. This observation is confirmed by the previous findings from the brainstem which also suggest that SRM is more dominant when the listing environments are noisiest (please refer to Chapter 3 of this thesis).

### 4.4.4. Cortical auditory evoked potential (CAEP)

Amplitudes and latencies of P1 and N1 components were analysed. Figure 46 to Figure 49 show CAEP P1 and N1 amplitudes and latencies versus SNR for spatially co-located and separated conditions.

For all measures of the CAEP components (P1 and N1 amplitudes and latencies), no significant main effect between test and retest was found on latency (Table 2), allowing test-retest averages of the waveforms.
4.4.4.1. P1 amplitude

Figure 46 shows P1 amplitude versus SNR for both spatial conditions.

4.4.4.1.1. Main effect of SRM on P1 amplitude

No significant effect of spatial condition was found on the amplitude of the P1 component (Table 2). These results suggest that spatial separation cannot be recorded electrophysiologically using the P1 amplitude.

4.4.4.1.2. Main effect of SNR on P1 amplitude

The main effect of stimulus SNR on P1 amplitude was significant (Table 2). With increasing SNR, P1 amplitude increased as well. Although based on the statistical analysis the effect of stimulus SNR is significant, it is difficult to suggest that an increase in stimulus SNR gives rise to an enhancement of the P1 amplitude. Although one can perceive this trend from 5 to 15 dB SNR, it becomes difficult to support this observation down to -5 dB SNR.

4.4.4.1.3. Main effect of target frequency paradigm on P1 amplitude

A significant difference was found between the two stimulus paradigms (Table 2), with the flat paradigm generating larger amplitudes than the downward staircase paradigm.

4.4.4.1.4. Interaction effects between SRM and SNR on P1 amplitude

No significant interaction effects between SRM and SNR were found on the amplitude of the P1 component (Table 2). This result suggests that if different SNRs would have different capacity to show SRM, pooling the data of different SNRs together offsets each other effects hence no SRM could be detected via P1 amplitude, if any.
4.4.1.5. Interaction effects between SRM and target frequency paradigm on P1 amplitude

The interaction effects between SRM and target frequency paradigm on the P1 component were significant suggesting that flat and staircase frequency paradigm could have different effect on P1 amplitude in spatial conditions (Table 2).

Figure 46. Means and standard errors of P1 amplitude in spatially co-located and separated conditions as a function of stimulus SNR, in a flat (left) and staircase (right) frequency paradigms.

4.4.2. P1 latency

Figure 47 shows P1 latency versus SNR for both spatial conditions.

4.4.2.1. Main effect of SRM on P1 latency

No significant effect of spatial condition was found on the latency of the P1 component while all data of different SNRs and test and retest were pooled together (Table 2).
4.4.4.2.2. Main effect of stimulus SNR on P1 latency

No significant effect of stimulus SNR was found on the latency of the P1 component (Table 2). This outcome was not expected. It suggests that any P1 latency results need to be interpreted carefully.

Figure 47. Means and standard errors of P1 latency in spatially co-located and separated conditions as a function of stimulus SNR, in a flat (left) and staircase (right) frequency paradigms

4.4.4.2.3. Main effect of target frequency paradigm on P1 latency

No significant difference was found between flat and staircase frequency paradigms (Table 2).

4.4.4.2.4. Interaction effects between SRM and SNR on P1 latency

No significant interaction effects between SRM and SNR were found on the latency of the P1 component (Table 2).

Figure 47 shows a visual separation between co-located and separated conditions at -5 dB SNR when visually inspecting the figure. Although it is acknowledged here that the reporting and the interpretation of these cortical results needs to be done very carefully, especially because of the noisy character of the recordings and the absence of any statistical interaction effects for this measure (Table 2), the consistent emergence of a significant effect of SRM at the lowest stimulus SNRs reported both in Chapter three (as reflected in
the main research question of this Chapter) and in the FFR section of this Chapter (Section 4.4.3), warrants a closer look at the -5 dB SNR condition. Moreover, the results from Section 4.4.1 suggest that at least at the behavioural level, there is a significant effect of SRM between the two different stimulus paradigms. For this reason, a planned comparison test demonstrated that there is a significant difference between co-located and separated conditions at -5 dB SNR for the staircase stimulus paradigm (p=0.027), but not for the flat stimulus paradigm (p = 0.64). This tentative result carefully suggests that spatial separation of distractors from the target facilitates the processing of the target signal in an adverse environment at the level of cortex, which is reflected in the reduction of the P₁ latency component.

To estimate the absolute benefit of SRM in dB, Figure 47 also depicts P₁ latency means and SDs as a function of different stimulus SNRs for the two spatial conditions. For the co-located condition, the means and standard errors of P1 latency at -5 and 15 dB SNR were 174.5±20.2 and 130.2±23.5 ms. For the separated condition, they were 136.2±21.1 and 119.0±22.5 ms, respectively. This corresponded to an average decrease (44.3 ms over 20 dB) of 2.22 ms per dB increase of the target SNR in the co-located condition. The decrease in latency at -5 dB SNR when spatially separating distractors from the target was 38.3 ms, or 17.3 dB in absolute benefit (38.3 ms divided by 2.22 ms per dB). Given the relatively high value of absolute benefit and the SRM difference not being significant, not much attention needs to be drawn to this number however.

For downward staircase stimuli (Figure 47), in the co-located condition, these latencies were 218.9±19.7 and 112.1±19.9 ms respectively and for the separated condition, they were 159.2±21.6 and 149.1±23.0 ms, respectively. This corresponded to an average decrease (106.8 ms over 20 dB) of 5.34 ms per dB increase of the target SNR in the co-located condition. The decrease in latency at -5 dB SNR when spatially separating distractors from target was 59.7 ms, or 11.2 dB in absolute benefit (59.7 ms divided by 5.34 ms per dB) with 95% confidence intervals of 1.3 and 18.8 dB.
4.4.4.2.5. Interaction effects between SRM and target frequency paradigm on P1 latency

The interaction effects between SRM and target frequency paradigm on the P1 component were not significant suggesting that flat and staircase frequency paradigm could not have different effect on P1 latency in spatial conditions (Table 2).

4.4.4.3. N1 amplitude

Figure 48 shows N1 amplitude versus SNR for both spatial conditions.

4.4.4.3.1. Main effect of SRM on N1 amplitude

No significant effect of spatial condition was found on the amplitude of the N1 component (Table 2). These results suggest that spatial separation cannot be recorded electrophysiologically using the N1 amplitude.

4.4.4.3.2. Main effect of stimulus SNR on N1 amplitude

No significant effect of SNR was found on the amplitude of the N1 component (Table 2). These results indicate that an increase in stimulus SNR did not result in an enhancement of the N1 amplitude, which is not what one generally would expect. Care hence must be taken when interpreting these results.

4.4.4.3.3. Main effect of target frequency paradigm on N1 amplitude

A significant difference between the two stimulus paradigms was found (Table 2), with the staircase paradigm generating larger (absolute) amplitudes than the flat paradigm.

4.4.4.3.4. Interaction effects between SRM and SNR on N1 amplitude

No significant interaction effect was found between SRM and stimulus SNR on N1 amplitude (Table 2).
4.4.4.3.5. Interaction effects between SRM and target frequency paradigm on N1 amplitude

The interaction effects between SRM and target frequency paradigm on the N1 component were significant suggesting that flat and staircase frequency paradigm could have different effect on N1 amplitude in spatial conditions (Table 2).

![Graph](image1)

Figure 48. Means and standard errors of N1 amplitude in spatially co-located and separated conditions as a function of stimulus SNRs, averaged across the test and re-test in a flat (left) and staircase (right) frequency paradigms.

4.4.4.4. N1 latency

Figure 49 shows N1 latency versus SNR for both spatial conditions.

4.4.4.4.1. Main effect of SRM on N1 latency

No significant effect of spatial condition was found on the latency of the N1 component when pooling all data together (Table 2).
4.4.4.4.2. Main effect of stimulus SNR on N1 latency

The main effect of stimulus SNR on N1 latency was significant (Table 2). With increasing SNR, N1 latency decreased correspondingly.

4.4.4.4.3. Main effect of target frequency paradigm on N1 latency

There was no significant difference between flat and staircase frequency paradigms (Table 2) suggesting that there is no reason to look at flat and staircase frequency paradigm separately.

4.4.4.4.4. Interaction effects between SRM and SNR on N1 latency

No significant interaction effects were found on the latency of the N1 component (Table 2).

Similar to the carefully interpreted results for P1 latency, a visually identifiable spatial separation can be noted at -5 dB SNR (Figure 49). Given the similarity between this result with other results at -5 dB SNR in Chapter 3 (a significant spatial separation effect at -5 dB SNR for FFR F0 amplitude) and the current chapter (a significant SRM effect at -5 dB SNR for FFR F0 amplitude and P1 latency), it was carefully decided to investigate -5 dB SNR conditions.
separately and to conduct a planned comparison again at this specific SNR. For the staircase paradigm, a significant effect of SRM \((F_{[1, 12]}=6.42, p=0.026)\) was found at -5 dB SNR with the co-located condition \((M= 326.5, SD=23.0)\) having longer N1 latencies than the separated \((M=274.2, SD=10.0)\) condition. For the flat stimulus paradigm, no significant effect of SRM was found at -5 dB SNR.

For the co-located condition, the means and standard errors of N1 latency at -5 and 15 dB SNR were 335.5±19.0 and 269.0±35.0ms and for the separated condition, they were 306.2±32.0 and 290.6±31.0ms, respectively. This corresponded to an average decrease (66.5 ms over 20 dB) of 3.33 ms per dB increase of the target SNR in co-located condition. The decrease in latency at -5 dB SNR when spatially separating distractors from target was 29.3 ms, or 8.8 dB in absolute benefit (29.3 ms divided by 3.33 ms per dB). Given the SRM difference, not being significant, not much attention needs to be drawn to this number however.

For downward staircase stimuli, these latencies were 326.6±23.0 and 273.8±35.0ms for the co-located condition, and for the separated condition, they were 274.2±31.0 and 311.2±28.0ms, respectively. This corresponded to an average decrease (52.8 ms over 20 dB) of 2.64 ms per dB increase of the target SNR in the co-located condition. The decrease in latency at -5 dB SNR when spatially separating distractors from target was 52.4 ms, or 19.9 dB in absolute benefit (52.4 ms divided by 2.64 ms per dB) with 95% confidence intervals of 2.8 and 36.9 dB.

**4.4.4.4.5. Interaction effects between SRM and target frequency paradigm on N1 latency**

The interaction effects between SRM and target frequency paradigm on the N1 component were not significant suggesting that flat and staircase frequency paradigm could not have different effect on N1 latency in spatial conditions (Table 2).
4.4.5. Correlation between behavioural LiSN-T and electrophysiological measures

Figure 50 and Figure 51 demonstrate the relationship between electrophysiological spatial advantage (measured using FFR amplitude (in nV), and ABR, P1 and N1 latencies (in ms), which are calculated by subtracting co-located and separated conditions, and behavioural spatial advantage of the LiSN-T (in dB) which is calculated by differential scores between co-located and separated speech perception thresholds (SRTs). No significant correlations were found between LiSN-T and FFR F0 amplitude (\( r=0.05, \ p=0.86 \)), ABR latency using the staircase paradigm (\( r=0.15, \ p=0.63 \)), P1 latency using the flat paradigm (\( r=0.23, \ p=0.45 \)), P1 latency using the staircase paradigm (\( r=0.19, \ p=0.087 \)) and N1 latency using the flat paradigm (\( r=0.07, \ p=0.81 \)). However, significant correlations were found between LiSN-T and ABR latency using the flat paradigm (\( r=0.64, \ p=0.02 \), expected direction) and N1 latency using the staircase paradigm (\( r=0.55, \ p=0.05 \), reverse direction). Given the multiple statistical tests that were conducted and the relatively high p-values in case of significance, it is difficult to place a lot of weight on these results.
Figure 50. Scatterplot of the correlation between LiSN-T and spatial advantage of FFR F0 amplitude at -5 dB SNR, ABR latency in flat and staircase frequency paradigms across all SNRs. The results of linear regression show no significant correlation between electrophysiological and behavioural spatial advantages except for the middle panel.
Figure 5.1. Scatterplot of the correlation between the LiSN-T and spatial advantage of P1 and N1 latency at -5 dB SNR in flat and staircase frequency paradigms. The results of linear regression show no significant correlation between electrophysiological and behavioural spatial advantages, except for the bottom right panel.
4.4.6. **Summary of the results**

Table 3 provides a summary of all significant results in terms of the required increase of stimulus SNR (in dB) to obtain the same effect of SRM when using the different objective waveform measure. It only displays the measures that provided a significant effect of SRM. The CAEP results are provisional (hence in brackets), and need to be confirmed in a follow-up study.

<table>
<thead>
<tr>
<th>Absolute benefit (dB) [95% confidence interval]</th>
<th>p-value</th>
<th>Restrictive conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABR wave V latency</td>
<td>13.8 [6.9 – 20.4]</td>
<td>0.0062</td>
</tr>
<tr>
<td>FFR F0 amplitude</td>
<td>4.3 [1.3 – 7.4]</td>
<td>0.0088</td>
</tr>
<tr>
<td>(CAEP P1 latency)</td>
<td>(11.2 [1.3 – 18.8])</td>
<td>(0.027)</td>
</tr>
<tr>
<td>(CAEP N1 latency)</td>
<td>(19.9 [2.8 – 36.9])</td>
<td>(0.026)</td>
</tr>
</tbody>
</table>

To recapitulate the results, although ABR wave V amplitudes do not seem to be a good candidate, latency might be a more sensitive indication to reflect SRM at the level of brainstem where the generators of wave V are situated. These observations are confirmed by the FFR, which suggest that SRM is represented at the level of the brainstem where the generators of FFR are situated, but only at the lowest SNR. This observation is confirmed by the previous findings from the brainstem in Chapter 3, which also suggest that SRM is more dominant when the listing environments are noisiest.

Although the CAEP recordings are considered noisy and the subsequent results difficult to interpret with a healthy dose of scepticism and caution, analysis of the P1 and N1 components suggests that spatially separating distractors away from the target stimuli has an electrophysiologically measurable influence but exclusively on latency, only at the lowest SNR measured, and only using the staircase frequency paradigm. This is a cautious interpretation given the lack of a significant effect of stimulus SNR and a SNR x SRM interaction effect on P1 and N1 latencies.
4.5. Discussion

The principal purpose of this study was to collect electrophysiological evidence of SRM in adult participants with normal hearing through simultaneous recording of ABRs, FFRs and CAEPs. The results in the first study (described in Chapter 3) showed that at least some neural processing with the potential to facilitate SRM occurs in the brainstem and can be measured electrophysiologically in the form of an increase of the amplitude of the FFR at the stimulus fundamental frequency. As the auditory cortex receives inputs from the same brainstem, it was hypothesised for this study that SRM results in amplitude augmentation and latency reduction for both brainstem and cortical responses when they are measured simultaneously. The discussion is divided into three parts, covering the ABR, FFR and CAEP results separately and discussing their implications. As reported in Table 3, significant differences between the co-located and separated conditions were found for ABR latency, at -5 dB SNR for FFR amplitude at the target stimulus F0, P1 latency (-5 dB SNR and staircase) and N1 latency (-5 dB SNR and staircase). The last two measures were treated with great caution due to poor waveform morphologies and are considered to be provisionally. These observations, in combination with the results from the first study in Chapter 3 indicated a significant effect of SRM for the FFR F0 amplitude at -5 dB SNR, and moved the focus on this specific SNR in the rest of this discussion. For ease of comparison, changes in nanovolt and millisecond values accompanying SRM are converted to equivalent changes in stimulus SNR, in dB (Table 3). This indicates how much the target SNR needs to be increased or decreased to obtain a similar increase in response amplitude or decrease in response latency when changing from a co-located condition to a separated spatial condition. For the different AEPs, the amplitude increase (in nV) or latency decrease (in ms) due to SRM at -5 dB SNR is divided by the improvement (in nV or ms per dB) derived in the result section, based on amplitude and latency values of the co-located condition only as this spatial setup was considered the reference condition.

Although it was not a hypothesis to find a correlation between behavioural and electrophysiological measures, no correlation was found between the different objective measures as an electrophysiological test and the LiSN-T as a behavioural test (Figure 50). This was somehow predictable for three reasons. First, because the FFR, ABR and CAEPs reflect the neural activities underlying spatial processing at the level of brainstem or the
auditory cortex while the behavioural measurement of spatial processing is a by-product of the entire auditory system associated with other centres such as sensory, cognitive and motor systems. Second, these objective measures and LiSN-T might possibly reflect different aspects of spatial processing. For example, as an interaction effect between attention and spatial separation in the FFR in Chapter 3 could not be found while for the LiSN-T (and the LiSN-S) it is a critical component of being able to accomplish the test successfully. Third, the LiSN-T has not been validated yet as a behavioural test, showing some inconsistencies outside the normative range (i.e., low absolute spatial advantages) what normally would not be expected of a normally hearing person without SPD (Figure 50). This indicates the LiSN-T still needs more scrutiny before it can be applied as a non-language specific spatial processing evaluation test.

4.5.1. ABR

The amplitude of ABR wave V was not significantly affected by SRM. These findings disagree with the results in both the first and the current study, in which moving competing sources away from the target stimulus resulted in a significant increase of FFR \( F_0 \) amplitude. Unlike the amplitude of FFR \( F_0 \), which can represent binaural processing in a rather straightforward way using the energy at the \( F_0 \) frequency, in ABR the more common index to express a neural representation of binaural processing is the binaural difference (BD) potential (or in some literature known as the binaural interaction component, or BIC) (e.g., Ballachanda and Moushegian, 2000, Brantberg et al., 1999, Clark et al., 1997, Furst et al., 1985, Gerull and Mrowinski, 1984, Riedel and Kollmeier, 2002). In the current study, due to limitations of the recording time, it was not possible to consider monaural along with binaural recordings to calculate this binaural index. Future study is needed to investigate whether BD represents binaural spatial processing mediated SRM.

Although the current experiment could not demonstrate a significant effect of spatial separation on ABR wave V amplitude, this observation can be explained in the light of previous research. Polyakov and Pratt (1996) and Pratt et al. (1997) suggested that ILD and ITD activities of binaural click ABR were processed independently because of their different effects on the ABR-BIC indicating that separate brainstem structures are involved.
in the processing of these spatial cues. Zhang and Boettcher (2008) in addition, showed different effects of ITD and ILD on the BIC of 80 Hz ASSR, which indicated that different neural populations in different pathways in the auditory brainstem encode the ITD and ILD. These findings suggest that binaural spatial cues are not combined to create an additive potential to represent auditory space in the early stages of the auditory system. Moreover, based on neurophysiologic findings, the result of the ABR wave V amplitude in the current experiment can also be explainable. It has been accepted that the SOC acts as the first location in the auditory pathways for binaural processing. The different neural populations in this nucleus are responsible for processing of binaural information and are the first centre for the calculation of differences in time and level (ITD and ILD) of the incoming signal (McPherson and Starr, 1993, Ungan and Yagcioglu, 2002). These cues still need to be more refined in the proceeding centres and amalgamated with monaural spatial information, which directly travel from the DCN into the IC neurons to merge with other spatial information coming at this level. Therefore, it seems possible that spatial information up to the IC, which is reflected in the ABR wave V amplitude, is still not fully organized to represent a comprehensive picture of a target sound location while in competition with distractors spatially. Taken together, these findings indicated that parameters of these spatial cues are not combined to create a larger spatial cue in early stages of the auditory system to differentiate the spatial location of a signal if the signal is interleaved with distractors. As a result, these cues could not be recorded appropriately via the far-field recording of the ABR wave V amplitude. Although the current experiment is not able to show which neural structure is responsible for the encoding of ILD and ITD cues, failing to detect these cues using the ABR amplitude may suggest that either these cues are not summed up together to create a recordable signal or cancel each other out. In addition, under the presence of a distractor, the ability to record these cues becomes worse.

A significant difference was found between flat and staircase frequency paradigms (Table 2), resulting in a larger wave V amplitude evoked by the flat paradigm when compared to the staircase paradigm. These differences were possibly caused by the extra variability (‘jitter’) introduced by the EEG data processing that corrected for the different delays introduced in the staircase paradigm. It is also possible the dissimilar fundamental
frequencies of the flat and staircase paradigms had an effect on wave V amplitudes. More research is needed to verify this explanation.

The latency of the ABR wave V reduced significantly when the competing sources in the horizontal plane were separated (at ±90⁰) from the target located at 0⁰ (Figure 42 and Figure 43 and Table 2). The separated condition was shorter in latency than the co-located condition, and achieved an absolute benefit of 13.8 dB (95% confidence interval between 6.9 and 20.4 dB). These numbers are not significantly different from the values reported by Cameron and Dillon (2007a) (12.9 dB), Allen et al. (2008) (10.7-12.8 dB), Brown et al. (2010) (12 dB), Cameron et al. (2011) (12.8 dB), but significantly different from Gelfand et al. (1988) (5-6 dB) in their behavioural counterparts. Some differences would be indeed expected because of differences in the specific targets and distractors though there is some spectro-temporal similarity. In this experiment, tone-complex stimuli were used both as targets and distractors whereas Cameron and Dillon (2011) and Gelfand et al. (1988) used two talkers and 2-speaker babble, respectively. In addition, only the spatial advantage up to the level of the brainstem was determined, which constitutes the preliminary stage of spatial processing.

The results also showed that as SNR increases, the ABR amplitude and latency increases and shortens, respectively. These results are roughly in agreement with the findings of the Burkard and Hecox (1983) study. Burkard and Hecox (1983) examined the effects of ongoing broadband noise at different levels ipsilaterally on ABR wave V using click stimuli. They demonstrated that the effect of different noise levels on a given click intensity resulted in a decrease of wave V latency. Based on Figure 2 (showing a 60 dB SPL target intensity curve) of the Burkard and Hecox (1983) study, it is calculated that the amount of reduction in latency they recorded was 0.6 ms per 20 dB SNR range (equal to a masking of 45 to 65 dB SPL which corresponded to 6.5-7.1 ms in latency). This is similar to the latency differences it is obtained in the present study (0.62 ms reduction in a 20 dB SNR range). Moreover, their observed effect of SNR on the ABR wave V amplitude is in accordance to the first experiment of this thesis investigating the effect of different SNRs on FFR F0 amplitude.

When different SNRs were tested in co-located and separated spatial conditions, no significant interaction effect was found between SRM and SNR for both ABR wave V
amplitude (p=0.18) and latency (p=0.067). This finding is in contrast to the first experiment reported in this thesis. In the first experiment, a significant interaction was found between SRM and SNR for FFR F0 amplitude.

The current data showed no significant interaction effect between SRM and stimulus paradigm on both amplitude and latency of wave V. To the author’s knowledge, the interaction effect of SRM and stimulus frequency paradigm on ABR wave V amplitude and latency has not been investigated to date. However, it has been psychoacoustically shown that when target and distractors are similar in context, any separation between the target and the distractors results in larger SRM.

Additional research is required to clarify the effects of SRM on both amplitude and latency of the ABR at different SNRs. Here are some recommendations regarding the use of the ABR to represent SRM in future research. First, future research may consider a larger sample size and particularly a greater number of stimulus presentations to increase the ABR quality. Second, future research may be aimed at finding the ABR BIC by considering both binaural and monaural recordings to investigate whether the BIC is correlated with SRM. And third, future research should consider using stimuli such as clicks that might better enable the effect of spatial separation on individual peaks to be determined. For example, wave III is thought to be generated in the SOC, which encompasses neural elements that are responsible for encoding of ITD and ILD cues.

4.5.2. FFR

The results of this part of the experiment yielded the following observations: a robust SRM is present at SNR -5 dB; the amplitude of the FFR increases systematically as a function of increasing SNR, and; there is a significant interaction effect of SNR and SRM.

No significant main effect of SRM on FFR F0 amplitude could be found. However, the interaction effect between SRM and SNR, which reflected via F0 amplitude, was significant. After a post-hoc test, a significant effect for SRM at -5 dB SNR was obtained. At this SNR, the amplitude of FFR F0 increased significantly when the distractors in the azimuth plane were separated (at ±90⁰) from the target located at 0⁰ azimuth. This observation is confirmed in the first study of this thesis (described in Chapter 3), where it is utilised speech elements as
target, and speech distractors as maskers at three different SNRs (-5, 0, and 5 dB). A significant effect of SRM was found here at -5 dB as well. These results are generally in agreement with the idea that introducing perceptual separation between distractors and targets result in binaural SRM in human psychoacoustic (Brungart et al., 2005, Cameron and Dillon, 2007a, Cameron and Dillon, 2008, Freyman et al., 1999, Huang et al., 2008, Huang et al., 2009, Rakerd et al., 2006) and animal electrophysiological (e.g., Caird et al., 1991, Du et al., 2009a, Du et al., 2009b, Grothe et al., 1996, Mandava et al., 1996, Yost and Sfeft, 1993) studies. A similar phenomenon to SRM is the binaural masking level difference (BMLD), which relies on interaural phase reversal of either signal or noise resulting in unmasking signal (Hirsh, 1948b). More specifically, release from masking due to the S0Nπ condition in the BMLD partly corresponds to the separated condition of the current experiment where all spatial cues are present however and the BMLD is solely based on ITD (or more precisely interaural phase difference). Signal unmasking in BMLD relies on the cross-correlation of signals in two ears and considers different binaural arrangements represented by different underlying neural populations. That is, when noise is correlated in both ears, any change in signal results in decorrelation of neural response to the signal (Müller, 2008, Palmer, 1997). In the current study, when shifting the distractors ±90° to the sides, two different interaural correlations occurred: one for the target and one for the distractors. Hence, two different neural populations likely contributed. This decorrelation can either release neural populations that are sensitive to the target signal coming from the 0° azimuth or alleviate the degrading effects of distractors by decreasing their overall energy level.

The enhanced amplitude of the spatially separated response, when compared to the co-located response in different findings of the current experiment, is thought to reflect complex neural interactions among binaurally spatial sensitive neurons, more specifically neural networks that innervate binaurally in the SOC (Boudreau and Tsuchitani, 1968, Guinan et al., 1972a, Guinan et al., 1972b, Moushegian et al., 1975) including MSO and LSO, and in the IC (Rose et al., 1966, Semple and Aitkin, 1979, Stillman, 1972). During proceeding of the target and the distractors being used in the current study towards the higher neural structures in the auditory cortex, the preliminary spatial processing of the target and the distractors occurs at the level of the SOC where neural units fuse binaural information. During the converging of information of the target signal, depending on the interaction of
Chapter 4: Investigating spatial processing through …

excitatory and inhibitory inputs coming from bilateral cochlear nuclei, spatial cues including ITDs and ILDs are encoded (e.g., Boudreau and Tsuchitani, 1968, Boudreau and Tsuchitani, 1970, Caird and Klinke, 1983, Guinan et al., 1972b, Joris and Yin, 1995, Kil et al., 1995). Encoding of ILDs and ITDs cues of the target signal occur in a close dynamic relationship with the ventral cochlear nucleus (VCN) via specific patterns of bifurcate activities of binaural neurons (Boudreau and Tsuchitani, 1970, Glendenning et al., 1985, Shneiderman and Henkel, 1985, Smith et al., 1993, Smith et al., 1998, Tolbert et al., 1982). In co-located condition of the current experiment, where both target and directors coming from the 0° azimuth, the inhibiting neural networks responsible for suppressing background noise decrease neural response to both target and distractors (Frisina et al., 1997). However, because of constant repeatability of the target over the time and randomization nature of the distractors, the target stimuli relatively becomes enhanced and hence relatively unaffected by inhibition networks and the distractors, in contrast, became suppressed (Frisina et al., 1997, Frisina et al., 1994). Then, this augmented target signal and lesser distractors will project towards the SOC centres for encoding ILDs and ITDs (Boudreau and Tsuchitani, 1970, Glendenning et al., 1985, Shneiderman and Henkel, 1985, Smith et al., 1993, Smith et al., 1998, Tolbert et al., 1982). However, the dominating encoding strategy of information is based on the target-to-distractors energy ratio particularly for the higher SNRs. When decreasing the energy level of the target, the vulnerability of the spatial cues will be increased accordingly and the spatial cue detectors encode both ILD and ITD cues of target and distractors (Du et al., 2009b). In fact, in this condition the target auditory stream is corrupted by distractors. In the separated condition, on the other hand, spatially removing the distractors from the target result in enhancing of the target energy, hence more accessibility of its spatial cues by the neural cue detectors. While encoded signals proceed through the auditory pathways towards the IC nuclei, they pass an extra level of refinement via neural populations in the lateral lemniscus (Yang and Pollak, 1994) while transgressing to the opposite side of the auditory pathways (Yang and Pollak, 1994). It is reasonable to assume that this transgression is for establishing another contrast of information arriving from both contralateral and ipsilateral auditory pathways and for converging this information with monaural information arriving at the level of the IC (i.e., spectral information of the monaural neurons and, ITD and ILD information of the binaural
neurons) (Oliver and Huerta, 1992, Schofield, 2005, Vater et al., 1995, Andersen et al., 1980). At the level of the IC, by separation of the distractors from the target spatially in the current study, some binaural sensitive neurons are induced exclusively by the target and not by the distractors, resulting in to an enhancement in the FFR SNR (Du et al., 2009b). Further investigating neural mechanisms underpinning the unmasking occurrence reflected in FFR, Du et al. (2009b) demonstrated that when there was a disparity between signals in the two ears of a rat with the ipsilateral signal leading the signal at the contralateral ear, both the contralateral dorsal nucleus of the LL and the neural population of the contralateral IC contribute to the unmasking occurrence reflected via the FFR. In fact, spatial separation of distractors from the target signal gives an opportunity to the auditory neural populations to suppress the distractors without affecting the target. This explanation could be employed as well for the spatial separated condition in the current experiment.

As demonstrated in Table 2, there is a significant effect of SNR on FFR F0 amplitude, which with increasing SNR, the amplitude of FFR F0 increased correspondingly. This finding is consistent with the finding in the first experiment of this thesis in Chapter 3, where it was found that with increasing SNR, FFR F0 amplitude to a speech vowel increased accordingly. This finding is also in accordance with the study by Billings et al. (2013) who used the speech syllable /ba/ in the presence of speech spectrum continuous noise and reported that as SNR increased, amplitudes became larger.

The absolute spatial advantage, converted from nV into dB, was equal to 4.3 dB (95% confidence interval of 1.3 to 7.4 dB) as shown in Table 3. This number is similar to the value was obtained in the first study of this thesis in Chapter 3, which was 3.3 dB (95% confidence interval of 0.3 to 6.3 dB). However, these values are significantly different from the values reported by Cameron et al. (2011) (12.8 dB) but consistent with the value that was reported by Gelfand et al. (1988) (5-6 dB) in their behavioural counterparts. This inconsistency between the spatial advantage found in the current experiment and the value that has been reported by Cameron et al. (2011) is possibly due to the difference in target and masker stimuli being used in this study. Alternatively, it might be because of the rather early stages of the auditory signal processing in the central auditory system where the FFR originates from.
It has been suggested that the most likely FFR generators are in the upper brainstem, presumably the IC (Smith et al., 1975). Hence, based on the results of this experiment and those reported in Chapter 3, and the assumed neurophysiology of SRM discussed above, it seems possible that the FFR might become an objective clinical tool with a good test-retest reliability for evaluating of SRM at the level of the brainstem. If it would be recorded sufficiently long (e.g. half an hour, but this value is arbitrarily chosen) for each spatial condition at a low SNR where SRM is measurable, as evidenced by the results in this and the previous chapter, an objective SRM score might be able to be calculated for each individual (and not just in a group analysis). However, further investigation in this field will be required to develop a clinical protocol to diagnose and monitor the SRM deficits in individuals with SPD and in individuals with SRM deficits but without SPD.

In combination with the results obtained from the ABR wave V latency, the hypothesis is supported that the FFR to tone-complex stimuli in a virtual simulation of the auditory space reflects a neural population that undertakes binaural SRM within the brainstem. It also suggests that the subcortical processing of auditory spatial events at the level of the brainstem is the starting point of disentangling target signals from distracting noise, at least for the acoustical conditions similar to the condition that has been utilised in this experiment.

4.5.3. CAEP

No significant effect of SRM and no interaction effect between SRM and SNR were found on P1 and N1 latencies. However, Figure 47 and Figure 49 (right panels) demonstrated a considerable visual separation between the two spatial conditions at -5 dB SNR while using the staircase frequency paradigm. As previously discussed, significant results were found in Chapter 3 (a significant spatial separation effect at -5 dB SNR for FFR F0 amplitude) and the current chapter (a significant SRM effect at -5 dB SNR for FFR F0 amplitude and a general effect of SRM with wave V latency). Therefore, it was considered admissible to look at -5 dB SNR separately and to conduct a planned comparison at this particular SNR for both P1 and N1 latencies. The latency of P1 and N1 components reduced significantly in the staircase paradigm only at -5 dB SNR when the distractors in the
horizontal plane were separated (at ±90°) from the target which located at azimuth 0°. This could suggest that P1 and N1 latency in the downward staircase frequency paradigm and at the lowest SNRs possibly could reflect SRM at the cortical level. The spatial advantages for the CAEPs (11.2 dB for P1 and 19.9 dB for N1) were higher than the absolute spatial advantage for the FFR results (3.3 and 4.3 dB respectively) in Chapters 3 and 4, but similar to the ABR results (13.8 dB) in Chapter 4 (Table 3). This might suggest, albeit very cautiously, that there is an additional processing step of SRM in the auditory cortex beyond the brainstem, at least when comparing the FFR results with the cortical results. However, because of the noisy cortical waveforms, especially at low SNRs, and given the fact that CAEPs are sensitive to stimulus SNR (Baltzell and Billings, 2014, Billings et al., 2013, Billings et al., 2007, Billings et al., 2009), one has to be cautious in interpreting these rather preliminary results. Nevertheless, it would be interesting to further investigate the effect of SRM at low SNRs at the cortical level.

Because of the target and the distractors similarity, the target and the distractors stimulate possibly similar groups of neurons in the auditory cortex. This may result in the activation of the neural groups that overlap together to a large extent. This overlapping may lead to a larger masking effect on activity of neural population underlying SRM. Unfortunately, this assumption could not be verified in the literature. To our knowledge, the effects of spatial co-location and separation on cortical auditory evoked potentials using flat and staircase tone-complex paradigms have not previously been investigated. Hence, it could not be found similar studies to compare our findings with them. Another possible explanation for the reduced latencies is that the combination of activity of left and right hemifield-tuned neural populations participated in response generation. This explanation has been suggested by several investigations on the monkey auditory cortex (Ahissar et al., 1992, Benson et al., 1981, Werner-Reiss and Groh, 2008, Woods et al., 2006). Based on the findings of these studies, it is possible in the current experiment that the single neurons in the auditory cortex encode an entire virtual hemifield - either the right or the left side or both sides for the small neural populations that are tuned to the frontal location only. However, the distractive effects of the maskers can still be present because the maskers were presented binaurally as well, but in different directions assuming that they might
occupy the activity of neural populations that are sensitive to the ±90° and not to the 0° directions.

In this experiment, blocks of flat and staircase stimuli were utilised. Flat stimuli had a single fundamental frequency (and its harmonics) whereas staircase stimuli had multiple fundamental frequencies (with their harmonics). The hypothesis of this study was that the use of these stimulus paradigms would result in different degrees of SRM, as evidenced in the behavioural component of this study (Section 4.4.1), where staircase stimuli resulted in larger spatial advantages. One of the possible explanations of this behavioural difference is that the distractor being used here affected flat stimuli only by creating energetic masking (EM) whereas for staircase stimuli it created both energetic and informational masking (IM) because of an increase in resemblance between targets and maskers in their frequency contents. As a result, spatial advantage increases when staircase stimuli are used instead of flat stimuli. It is well documented that when target and masker have similar spectrotemporal content such as speech masked by (time-reversed) speech, EM is not the dominant masking factor (Brungart and Simpson, 2002, Culling et al., 2004). If target and masker are easily confused with each other, differences in the perceived locations of the masker and target cause an increase in spatial unmasking (Durlach et al., 2003a, Durlach et al., 2003b) or SRM which is typically known as spatial unmasking from IM. Although significant interaction effects were found between the stimulus paradigm and SRM for P1 and N1 amplitudes, and larger responses were identified for the flat stimulus paradigm for wave V, P1 and N1 amplitudes (Table 2), it is difficult to find any evidence for an electrophysiological correlate of the behavioural difference between stimulus paradigms.

A significant main effect of SNR was only found for P1 amplitude and N1 latency and not in other conditions, which are likely because of the noisiness of the recorded waveforms (as mentioned earlier). The decrease in latency with increasing SNR that has been found in the current experiment is in agreement with those studies that have reported that latency generally reduces with increasing SNR (Baltzell and Billings, 2014, Billings et al., 2013, Billings et al., 2007, Billings et al., 2009). Moreover, increasing amplitude with increasing SNR in the current study is in accordance with the studies that have been conducted on cortical neurons via a near-field investigation which they have showed that with reducing noise levels, amplitude responses enhance proportionally (Phillips, 1985, Phillips, 1990,
Phillips and Cynader, 1985, Phillips and Hall, 1986). When compared to the psychoacoustical investigations, the current results are in agreement with these studies as well which they have demonstrated that the decrease of masking level is proportional to the increase in detection of signal threshold. This dynamic interaction between signal and noise can also be tracked through the auditory cortex (Hawkins Jr and Stevens, 1950, Stevens and Guirao, 1967).

When stimulus paradigms were tested for both P1 and N1 components in this experiment, the data showed a significant effect of flat and staircase stimulus frequency on amplitudes of these cortical components. However, this effect was not significant for the latency of the components. Because larger responses were identified for the flat stimulus paradigm (P1 amplitude) and for the staircase stimulus (N1 amplitude) (Table 2), no conclusion can be put forward regarding the effect of stimulus type on amplitudes of the cortical components.

4.6. Conclusion

This study set out to investigate the effects of SRM at different SNRs on electrophysiological parameters including amplitude of the FFR, and amplitude and latency of the ABR and the CAEP, recorded in participants with normal hearing. It was revealed that an electrophysiological representation of SRM could be recorded via ABR latency, only at the lowest tested SNR of -5 dB for FFR F0 amplitude (which confirmed the results obtained in Chapter 3), and via P1 and N1 latencies (using the staircase frequency paradigm), but only at -5 dB SNR as well. The observations at the cortical level are premature at best due to noisy data, and need to be confirmed in a follow-up study. The main research question hence can be positively answered: there seem to be evidenced that some neural population activities in the brainstem, and possibly in the auditory cortex might underpin SRM. The results of Chapter 3 involving the FFR at low SNRs could be reproduced indeed. SRM indeed seems to decrease ABR latency, but not ABR amplitude. A careful suggestion could be identified in the data that might indicate that CAEP latency also decreases at low SNRs, but only for the staircase paradigm. The results in this study validates that the central auditory nervous system at the level of the generators of these components utilises spatial cues for separating target from any distractors, but mainly at lower SNRs. This confirms that the
central auditory system has the capability to filter out irrelevant background noise on the basis of spatial information, and that these effects are larger in the more challenging listening environments than less challenging environments.

Two minor research questions could be answered as well. First, there is no clear evidence that the design of the target stimuli affects SRM processing in the brainstem and cortex. It does however in the behavioural correlate that it was investigated. The tentative results for the cortical latencies and the staircase stimulus paradigm might suggest that at the cortical level a larger effect of SRM can be established by using a staircase paradigm. This is not backed up by any statistics however, and more research is required. Second, this study showed it is possible to simultaneously record ABR, FFR and CAEPs, as has been confirmed earlier by (Lauffer et al., 1993). The CAEPs however were of very poor quality, which makes it necessary to change future research protocols.

Given the positive results obtained by the ABR latency, FFR F0 amplitude and - to a less convincing extent - the latency of the P1 and N1 cortical components, the used objective tools could be suggested for the evaluation of the spatial processing at the level of the brainstem (and the auditory cortex). This study suggests that using tone-complexes in a block arrangement as a target and distractors may be useful to measure SRM in the brainstem. Using these stimuli to evaluate SRM in the auditory cortex is a matter of debate however. Before the procedures that have been used in this experiment can be employed clinically for evaluation of spatial processing and possible spatial processing disorder at the level of the brainstem, validation of this method is required for evaluating adults with SPD, normal children and children with SPD, and elderly both normal and with SPD. An issue it has not be tackled is the extent to which the measurements are sufficiently sensitive to be meaningful when applied to a single individual. Moreover, additional research is needed to determine what other neural structures at the level of the brainstem from the SOC up to the LL and the IC through the evaluation of different components of the ABR (like wave III), at the level of the midbrain using middle latency responses, and beyond the CAEP generators using late-latency event related potentials like the mismatched negativity (MMN), P300, and N400 are involved in auditory spatial processing.
Chapter 5. Summary, limitations, clinical significance, and future research
5.1. Preface

In this chapter, the first two chapters and the results of the two experiments reported in Chapters 3 and 4 are summarised and discussed. Final conclusions are derived, and future directions for research are proposed based on the findings in this thesis.

5.2. Summary

5.2.1. Introduction

The impetus for this research was partly prompted by the need to find objective SRM testing as an alternative to subjective SRM testing. In current use are Listening in Spatialised Noise – Speech (LiSN-S) and Listening in Spatialised Noise – Tonal (LiSN-T) tests, based on speech and tone complexes respectively, the latter still being in development. Both tests provide behavioural measurements of SRM, and LiSN-S has been shown to be sensitive to deficits in spatial release from masking (SRM) in different populations beyond the age of 6.

A behavioural test, which is a subjective method, requires the examinee to respond to the presented stimuli appropriately. This is generally difficult or impossible in young children. In contrast utilising an electrophysiological test, which is an objective method, does not rely as much on subject participation. In this thesis, using an electrophysiological approach, the effect of spatial separation on the detection of sounds in a competing acoustic environment was investigated by examining the electrophysiological responses recorded from the auditory brainstem and auditory cortex. Two questions were addressed. First, do electrophysiological measures of neural responses to auditory input provide evidence of processing underpinning SRM in the auditory system? Second, do electrophysiological measures of neural responses to auditory input identify processing underpinning SRM in either the brainstem or cortex (or both)?
5.2.2. Neurophysiological and electrophysiological approaches to spatial processing: theoretical bases

Chapter 2 provided an overview of the literature, which suggests that those with peripheral normal hearing but with central auditory processing disorder ((C)APD) appear to struggle with understanding target messages in adverse listening environments, where there are different sources of distractors interfering with each other and with the target signals. Individuals suffering from ((C)APD) are unable to access those cues that play essential roles in disentangling target speech from distractors. A subcategory of this diverse population are those with spatial processing disorder (SPD) whom are specifically unable to attend a target signal and ignore distractors, while both targets and distractors are arriving from different angular locations in an auditory space.

Understanding SPD in the central auditory system relies on our understanding of how the central auditory mechanisms that underlie spatial processing operate in normal-hearing individuals. In a normal-hearing population, the processing of a signal of interest intermingled with noise is a complex process that starts at the auditory nerve and moves upward to higher neural structures. It seems likely that two refinement processing stages are at work in the central auditory nervous system (Yost and Gourevitch, 1987): monaural information processing which is based on spectral (Wightman and Kistler, 1997) and distance cues (Zahorik et al., 2005), and binaural information processing which is dependent on interaural intensity level difference (ILD) and interaural time difference (ITD) cues (Middlebrooks and Green, 1991).

Decoding of the monaural and binaural cues of a target signal at the level of the brainstem depends on preliminary processing of target and distractors at the level of the auditory nerve fibres. The dynamic range of most auditory nerve fibres is changed in the presence of distractor noise in the way that the distractor increases both the nerve’s threshold and the level needed to reach the maximum firing rate (Costalupes et al., 1984, Geisler and Sinex, 1980, Gibson et al., 1985). The preliminary processed information then arrives at the cochlear nuclei (CN) where the target–distractor noise ratio (essentially a signal-to-noise ratio or SNR) improves. This improvement in SNR occurs by dendritic filtering and postsynaptic membrane activities. As a result synchronous responses to signal, and
inhibitory inputs increase resulting in lowering of the average responses to signal and noise (Frisina et al., 1997). Hence, the averaging of random noise in combination with the synchronisation of the signal amplifies the signal in comparison to the noise.

Further improvement in SNR occurs via an adjusting neural loop between the superior olivary complex (SOC) and the ventral cochlear nucleus (VCN), which optimises the input from the VCN to the SOC. This loop enables the lateral superior olive (LSO) to extract ILD cues and the medial superior olive (MSO) to extract ITD cues (Boudreau and Tsuchitani, 1970, Glendenning et al., 1985, Shneiderman and Henkel, 1985, Smith et al., 1993, Smith et al., 1998, Tolbert et al., 1982). ILDs and ITDs associated with other signal features, while receiving more refinement at the lateral lemniscus (LL) level, will project to the inferior colliculus (IC) and converge into a single node. The output of the IC then proceeds towards the medial geniculate body (MGB) to receive further processing on monaural spectral information. Finally, these signals are delivered to the auditory cortex for determination of the sound source location.

Although both ILD and ITD cues can each enable source determination, ILD cues are more likely to be critical in spatial processing (Glyde et al., 2013b, Kidd et al., 2010). This finding was supported by Phillips and Irvine (1981) who reported that a population of neurons exists which are most sensitive to signals with zero ILDs and hence show maximal excitation to binaural stimuli at or near 0° azimuth. This insight may explain why humans preferentially locate all intended sources within the frontal plane. For example, from childhood, we learn that when we hear an interesting sound, we turn toward it and try to locate its source by looking directly at it. Repetition of this behaviour over time may have strengthened the neural response underlying the encoding of frontal source location.

While spatial processing has been explored in both animal and human studies evaluating spatial cues (in azimuth and elevation) in lateralisation and/or localisation experiments, a methodological sequence of spatial processing under a naturalistic environment has yet to be established. To identify how target stimuli in the presence of distractors under a natural environment are processed in the auditory system, auditory evoked potentials (AEPs) can be employed. AEPs measure electrical activities of neural populations in the auditory system in response to appropriate stimuli. The AEPs it was utilised in this thesis were the auditory brainstem response (ABR), the frequency following
response (FFR), and the cortical auditory evoked potential (CAEP). Based on the objective and research questions of this thesis, two studies were conducted, which were explained in Chapters 3 and 4.

5.2.3. Spatial release from masking: Human frequency-following response correlation

Chapter 3 described the first study addressing the question whether electrophysiological tests can be used to detect SRM. The first hypothesis was that a significant effect of SRM could be measured through the amplitude of the FFR, an objective measure, which is recorded in the brainstem in response to a speech stimulus. The second hypothesis was that a neural representation of SRM was modulated by attention. Adult participants with normal hearing were evaluated using the FFR in response to a speech stimulus /u/ in two spatial conditions, i.e. co-located and separated, masked by two talker-distractors. Three SNRs (-5, 0 and 5 dB) were considered, in combination with an attention and a non-attention condition.

A significant SRM was found at the lowest SNR, as evidenced by an increase in FFR F0 amplitude in the separated condition compared to the co-located condition. This suggests that when distractors are shifted to the sides, the neural populations responsible for SRM were able to use the spatial differences between the target and distractors to increase the neural representation of the frontal target. Unexpectedly, there was no SRM at the higher SNRs (0 and 5 dB). It is possible that when the target signal is equal to or higher in intensity than the masker, the auditory system for processing of target stimuli, at least at the level of the brainstem, does not need SRM to adequately analyse the target.

Another finding of this study was that attending to the target stimuli resulted in enhancement of the FFR F0 amplitude when compared with ignoring the target stimuli. This result suggests that attention can modulate FFR responses at the level of the brainstem. Also unexpectedly, there was no significant interaction between SRM and attention. This result suggests, at least for these stimuli, either that frontal sounds are emphasised relative to sounds from other directions no more strongly when attended than when not attended,
or that the effects of attention on spatial processing occur in parts of the auditory system that are not involved in the generation of the FFR, for example in the auditory cortex.

Table 4 presents the electrophysiological spatial advantage in dB, indicating how much the target SNR in the co-located reference condition would need to be increased to obtain the same increase in response amplitude that occurs when changing from a co-located to a separated spatial condition. The amplitude of the FFR at \( F_0 = 110 \) Hz shows a spatial advantage of 3.4 dB and 3.3 dB for non-attention and attention, respectively. Thus, based on the results of the first study, the hypothesis could be supported that electrophysiological tests can be used to detect SRM, more specifically in the brainstem at low SNRs. The small size of the SRM compared to the SRM measured behaviourally, especially given the lack of effect of attention on SRM, leaves open the possibility that additional SRM occurs higher in the auditory system.

5.2.4. **Investigating spatial processing through the simultaneous recording of the auditory brainstem response, frequency following response, and cortical auditory evoked potential**

Chapter 4 reported on the second study, addressing the second question of this thesis, which hypothesised that neural responses underpinning SRM could be measured in the brainstem, or cortex, or both. This study also investigated whether brainstem and cortical evoked potentials could be recorded simultaneously. In this experiment a series of tone-complexes were presented which were developed based on the LISN-T paradigm used by Buchholz et al. (2013). The stimuli included two frequency paradigms: flat and staircase. The flat frequency stimuli had a fundamental frequency \( F_0 \) of 325 Hz, and harmonics up to 6 kHz. The staircase stimuli were composed of \( F_0 \)s ranging from 475 down to 175 Hz, and harmonics up to 6 kHz. In contrast, the masker was constructed using \( F_0 \)s randomly distributed from 100 to 550 Hz with harmonics up to 6 kHz. Blocks, each containing 11 tone complexes as target and masker, were presented. Adult participants with normal hearing were assessed behaviourally using the LISN-T, and electrophysiologically via ABR, FFR, and CAEP in two spatial conditions, five SNRs, test, and re-test conditions. Results demonstrated
that SRM can be recorded objectively both at the level of the brainstem, via ABR latency (an equivalent spatial advantage of 13.8 dB) and FFR amplitude (investigated for the flat paradigm only, 4.3 dB equivalent improvement at -5 dB SNR), and the auditory cortex (P1 and N1 latencies, significant for the staircase paradigm only, 11.2 and 19.9 dB improvement at -5 dB SNR, respectively). However, the results of this study at the cortical level need more cautious interpretation due to noisy data, and hence wide confidence intervals in the equivalent SRM, and need to be confirmed in a follow-up study. Further, our findings confirmed the observations in the first study (Chapter 3) that SRM could be observed only at lower SNRs for the FFR F0 amplitude.

Table 4 summarises the spatial advantages that are obtained using SRM for the different tested conditions. The similarity of benefit for the latencies of the different objective measures in the brainstem and cortex, when expressed as the equivalent change in input SNR, suggests that the spatial processing it has been measured predominantly occurred in the brainstem, the effects of which could then also be seen in the auditory cortex. Therefore, when considering the hypotheses of this study, our findings support the notion that SRM occurs primarily in the brainstem and progresses towards the auditory cortex. These benefits are similar to what is obtained through its behavioural counterpart, the LISN-S and the LISN-T (as shown in Table 4). When considering the amplitudes of the objective measures, only the FFR F0 amplitude comes out to be significant (in contrast with non-significant effects of SRM on ABR and CAEP amplitudes), with benefits that are much smaller than AEP latencies and behavioural measures. It is unknown why this is the case and further research is warranted. Further, it has been feasible to record both brainstem and cortical evoked potentials simultaneously using different protocols in the literature already (Bidelman, 2015, Lauffer et al., 1993). Lauffer et al. (1993), using pure tones interleaved with clicks, reported increased latencies and decreased amplitudes of N1 and P2. Bidelman (2015), applying clustered stimulus presentation and variable ISIs with fast and slow rates, showed that both brainstem and cortical responses could be recorded, with same amplitudes and a 3-fold increase in recording efficiency because habituation was reduced. In the current study, it was used blocks incorporating tone complexes with the tone complex onsets evoking brainstem responses (ABR and FFR) and the block onsets evoking CAEPs, allowing us to simultaneously investigate SRM at the brainstem and the cortex level.
ABR latency, FFR F0 amplitude and P1 & N1 latencies could qualify as indexes for the objective evaluation of spatial advantage. Nevertheless, due to the poor quality of the CAEP waveforms more research is required on the effects of SRM on CAEPs.

5.3. Objective and research questions

The objective of this thesis was:

To investigate the effect of spatial separation on the detection of sounds in a competing acoustic environment, by examining the electrophysiological responses recorded from the auditory brainstem and auditory cortex using standard electrophysiological measures.

Based on this objective, the following research questions were considered:

1. **Firstly, do electrophysiological measures of neural responses to auditory input provide evidence of processing underpinning SRM in the auditory system?**
   Study 1 in Chapter 3 indeed provided evidence that SRM could be objectively detected in the brainstem using the amplitude of the fundamental frequency of the FFR. This finding was confirmed by Study 2 in Chapter 4 at the level of the brainstem, while adding additional evidence through ABR latency, and tentatively extended towards the auditory cortex using CAEP latencies.

2. **Secondly, do electrophysiological measures of neural responses to auditory input identify processing underpinning SRM in either the brainstem or cortex (or both)?**
   Results in the first and second study suggest that processing underpinning SRM in both the brainstem and the cortex can be identified using objective measures like the ABR, FFR and CAEP.

5.4. Limitations

Although this dissertation has investigated spatial processing using objective approaches in participants with normal hearing, there are limitations that should be noted. The first study used the speech element stimulus (/u/) as the target and two speech streams as the distractors. The second study used tone-complex stimuli for both the target and the distractors. The stimuli for the first study were inspired by the LiSN-S (Cameron and Dillon,
2008) and for the second study were based on those used with the LiSN-T (Buchholz et al., 2013). The LiSN-S test employs complete sentences as target and running speech streams as the distractors and, as a result, is high in informational masking. In our study, by contrast, it was used a single synthetic phoneme, /u/, which because of its dissimilarity to running speech, was presumably more affected by energetic than informational masking, even though both target and distractors were speech-based stimuli. Because of its lower informational masking, it might be expected less SRM than is measured in the behavioural tasks.

In the second study, it was used tone-complex stimuli similar to those in the LiSN-T paradigm, both in a flat and a staircase paradigm. Buchholz et al. (2013) indicated that the texture of target and its similarity with distractors in a staircase approach creates informational masking, while a flat frequency paradigm does not. Although the frequency composition as is used in the LiSN-T was not exactly replicated, the LiSN-T uses a F0 range of 200-500 Hz with the distractors’ F0 ranging from 100-600 Hz, versus 475-175 Hz for target and 100-550 Hz for maskers in the second study, it seems likely that at least the results derived from the staircase paradigm match the behavioural results in Table 4. Furthermore, the large confidence interval surrounding this estimate shown in Table 4, is likely the result of the noisiness of the cortical recordings.

In the first study, the factor of attention was considered assuming that the top-down neural mechanisms can modulate the activities of the neural populations underlying spatial processing. As no significant interaction between attention and spatial separation in the brainstem was found, and as the test became too demanding for the participants to keep their concentration on the required task, it was decided to not include the factor of attention in our second study. Apart from the challenges encountered when trying to keep the attention of the participants, several technical issues have been identified when considering inclusion of attention into the second experiment of this thesis. First, adding a factor of attention would have resulted in additional noise in the data hence affecting the response SNR negatively (which was rather low already given the low stimulus SNRs in the second study). Second, adding a deviant target stimulus would have increased recording times, which were considerate in the second study already: if the factor of attention were to
Table 4. Statistically significant spatial advantages (in dB) of FFR F0 amplitude, ABR wave V latency, CAEP N1 and P1 latency, and behavioural tests LiSN-S and LiSN-T. Values were obtained at -7.65 dB SNR (average SNR) for the ‘low SNR’ condition in Study 1, and at -5 dB SNR for FFR and CAEP in Study 2. For ABR in Study 2, values were obtained over all SNRs. Calculations of the effect of SNR on the electrophysiological response were based on an SNR range of 10 dB in Study 1 and a range of 20 dB SNR in Study 2. Values between brackets for the cortex results are provisional – although significant at -5 dB SNR based on a planned comparison test given the results already found in study 1– as the CAEP waveforms were considered noisy and there was no significant interaction effect found between SRM and SNR warranting further analysis on separate SNRs.

<table>
<thead>
<tr>
<th>Study</th>
<th>Assessment tool</th>
<th>Stimulus type, paradigm and SNR</th>
<th>Non-attention (dB)</th>
<th>Attention (dB)</th>
<th>95% Confidence Intervals (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Brainstem</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>FFR Amplitude</td>
<td>Speech Low SNR (-7.65 dB)</td>
<td>3.4</td>
<td>3.3</td>
<td>0.31 - 6.3</td>
</tr>
<tr>
<td>2</td>
<td>FFR Amplitude</td>
<td>Tone Complex Flat -5 dB SNR</td>
<td>4.3</td>
<td>-</td>
<td>1.3 - 7.4</td>
</tr>
<tr>
<td>2</td>
<td>ABR Latency</td>
<td>Tone Complex Flat and staircase</td>
<td>13.8</td>
<td>-</td>
<td>6.9 - 20.4</td>
</tr>
<tr>
<td><strong>Cortex</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>P1 Latency</td>
<td>Tone Complex Staircase -5 dB SNR</td>
<td>(11.2)</td>
<td>-</td>
<td>(1.3 - 18.8)</td>
</tr>
<tr>
<td>2</td>
<td>N1 Latency</td>
<td>Tone Complex Staircase -5 dB SNR</td>
<td>(19.9)</td>
<td>-</td>
<td>(2.8 - 36.9)</td>
</tr>
<tr>
<td><strong>Behavioural</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>LiSN-T</td>
<td>Tone Complex Flat</td>
<td>-</td>
<td>10.7</td>
<td>6.0 – 15.4</td>
</tr>
<tr>
<td>2</td>
<td>LiSN-T</td>
<td>Tone Complex Staircase</td>
<td>-</td>
<td>14.0</td>
<td>10.4 – 17.7</td>
</tr>
<tr>
<td>Literature</td>
<td></td>
<td></td>
<td>-</td>
<td>12-14</td>
<td>-</td>
</tr>
</tbody>
</table>

164
be included, the assessment session would have lasted about 8 hours. This would have been unacceptable. However, in retrospect, it would have been interesting to have investigated the effects of attention on CAEPs as well. Therefore, it is highly recommended that the factor of attention is included in future experiments to determine whether attention affects spatial processing at the level of the auditory cortex and beyond.

Although a significant effect of SRM in some measures was found, the effect for other measures (e.g. ABR and CAEP amplitude) was not significant. It is possible that any non-significant effects found in the study were due to the limited number of subjects that participated in both experiments, especially because spatial processing appears to have its major effects at lower SNRs, and at these SNRs, the evoked responses are more affected by background electrophysiological noise. Further, tone complexes were used with a short interstimulus interval (ISI) in experiment 2. As a result, both the latency and amplitude of these responses were influenced in an unusual way by the overlapping CAEPs in response to tone complex onsets. Future research should focus on longer recordings in an individual participant, and/or more participants, especially at lower SNRs and while considering less complex stimuli.

5.5. **Clinical significance**

This dissertation demonstrated that SRM can be recorded objectively via both cortical and subcortical evoked potentials. People with normal hearing showed an increase of FFR amplitude and a decrease of ABR and CAEP latency in response to a target stimulus that is spatially separated from its maskers. If it can be shown that the electrophysiological representation of SRM is smaller in people with spatial processing disorder (SPD), it might be possible to develop an objective test for SPD (similar to its behavioural equivalents, the LiSN-S and LiSN-T) which would allow the evaluation of young children or those who are difficult to test, both of whom cannot provide reliable behavioural feedback. An obvious difficulty is that it was only just able to show significance for the effects of spatial separation in a group of 18 (Chapter 3) and 13 (Chapter 4) participants. Electrophysiological testing using the same set of conditions it has been used would clearly not be sufficiently sensitive when applied to an individual, especially as the individual may have only a reduced
effectiveness of SRM rather than the complete absence of ability to benefit from spatial separation. None-the-less, it may be possible to focus the available testing time on just one or two conditions of greatest interest, and so provide more reliable measures of evoked response strength than we were able to when the testing time was spread across multiple conditions.

5.6. **Directions for future research**

Auditory evoked potentials can play an important role in diagnosis and monitoring of central auditory processing disorders, more specifically spatial processing disorder. The use of AEPs to simultaneously evaluate various levels of the central auditory nervous system affords the potential to track the processing of an auditory signal from the periphery through to the more central parts of the auditory system. This continuum of real-time-monitoring of acoustical events could afford clinicians and researchers an opportunity to detect neurological obstacles to normal signal processing up to the regions responsible for signal comprehension. Although Bidelman (2015) has reported an optimal paradigm for recording cortical and brainstem responses simultaneously, the current study suggests an alternative approach to record cortical and subcortical neural activities in a short time (about 4 minutes). However, the observed CAEPs were noisy and highly unusual in their morphology and latency, and it is possible that this is the result of the multiple onsets created by the use of a series of tone complexes. Each tone complex created an individual cortical response, which overlapped each other and resulted in increasing latency and changing amplitude, hence deforming the morphology of the cortical waveform. To overcome this problem, future research should consider either a larger number of epochs or changing the stimuli.

Future research could investigate:

- whether it is possible to detect spatial processing at an earlier level, e.g. the SOC via the wave III of the ABR;
- what is the role of ITD, ILD and spectral cues separately and in combination in spatial processing of target-competitor complexes;
• the effects of different azimuth and elevation degrees on spatial processing of target-competitor complexes;
• the effects of shifting the target signal to the side while keeping the distractors in the frontal plane on spatial processing of target-competitor complexes;
• the effects of distance, reverberations, echoes, the precedence effect, and the recency effect on spatial processing of target-competitor complexes. The precedence effect refers to the first wave front in an echoic condition. If a sound is followed by another sound that is separated by a short time delay, people perceive a single fused auditory image. The perceived spatial location is dominated by the location of the first-arriving sound. In the recency effect, when a chain of acoustic stimuli proceeds through the auditory system, the last part of the information is more salient than the earlier part. Hence, the later part is more likely to contribute towards spatial separation than the earlier part of the signal;
• the attentional effects of top-down neural regulation and whether it is involved in background noise management;
• the role of short-term and long-term memory in spatial processing of target-competitor complexes. In some people with (C)APD, a relationship between SPD and deficient memory can be found;
• the effects of age (from childhood to adulthood and beyond) on spatial processing of target-competitor complexes;
• how the after-effects of (chronic) otitis media affects the neural populations underlying processing of target-competitor complexes;
• how noise-induced hearing loss might influence spatial processing of target-competitor complexes;
• whether the effects of spatial processing on CAEPs are different when more conventional (single onset) stimuli are used to generate the CAEPs, rather than the multiple-onset stimuli that it was used to generate CAEPs and ABRs from the same stimulus.
5.7. **Conclusion**

The aim of this dissertation was to investigate whether an index of SRM could be identified using electrophysiological responses and, using this index, to investigate whether the neural populations undertaking spatial processing were located in the brainstem or the cortex, or both. Employing ABRs, FFRs, and CAEPs in response to speech and tone complex stimuli in the presence of co-located and separated distracting stimuli of a similar nature at different SNRs, SRM ability was found to be present at the lowest SNR tested. This result suggests that the central auditory mechanisms objectively measured at the level of the brainstem and the cortex utilise spatial cues to separate signal from distractors mainly at lower SNRs. These SNRs correspond to conditions often encountered in real-life environments, and of course are the situations where the SNR enhancement provided by spatial processing is most needed. The conclusions of this dissertation possibly pave the way to electrophysiologically evaluate spatial processing disorder via FFR amplitude, ABR and CAEPs latencies subject to confirming the results of this project, at least at the level of the cortex in future studies. Further investigations in different populations including infants, children, and (elderly) adults are warranted to further develop the methods used in this thesis into a test suitable for clinical practice.
References


References:


References:


References:


DIAMOND, M. E., ARMSTRONG-JAMES, M., BUDWAY, M. J. & EBNER, F. F. 1992. Somatic sensory responses in the rostral sector of the posterior group (POm) and in the ventral posterior medial nucleus (VPM) of the rat thalamus: dependence on the barrel field cortex. *Journal of Comparative Neurology,* 319, 66-84.


References:


FREYMAN, R. L., BALAKRISHNAN, U. & HELFER, K. Release from informational masking in speech recognition. MidWinter Meeting of the Association for Research in Otolaryngology, 2000 St. Petersburg Beach, FL, USA.


References:


References:


References:


References:

References:


References:


References:


References:


References:


References:


References:


References:


References:


References:


Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Rouhbakhsh, Nematollah

Title:
Investigating the effect of spatial separation on the detection of sounds in competition, by examining electrophysiological responses from the brainstem and auditory cortex

Date:
2016

Persistent Link:
http://hdl.handle.net/11343/92484

File Description:
Investigating the effect of spatial separation on the detection of sounds in competition, by examining electrophysiological responses from the brainstem and auditory cortex