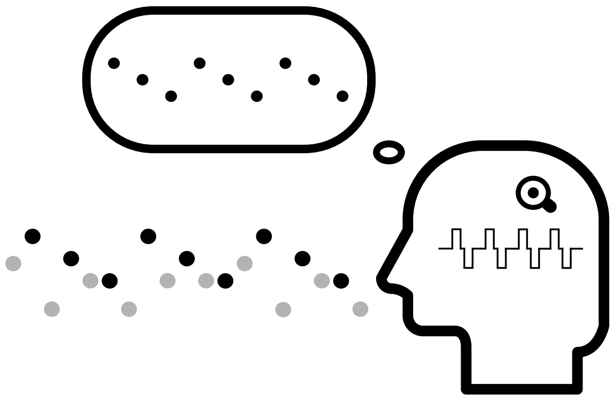
CONTRIBUTIONS TO
HEARING RESEARCH

Volume 35

Andreu Paredes Gallardo

Behavioral and objective measures of auditory stream segregation in cochlear implant users



DTU Electrical Engineering
Department of Electrical Engineering

Behavioral and objective measures of auditory stream segregation in cochlear implant users

PhD thesis by

Andreu Paredes Gallardo

Preliminary version: May 28, 2018

HEARING SYSTEMS

Technical University of Denmark

2018

This PhD dissertation is the result of a research project carried out at the Hearing Systems Group, Department of Electrical Engineering, Technical University of Denmark (Kgs. Lyngby, Denmark). Part of the project was carried out at The Bionics Institute of Australia (Melbourne, Australia).

The project was partly financed by the Oticon Centre of Excellence for Hearing and Speech Sciences (2/3) and by the Technical University of Denmark (1/3). Part of the research equipment was provided by Cochlear Ltd.

Supervisors

Main supervisor

Assoc. Prof. Jeremy Marozeau
Hearing Systems Group
Department of Electrical Engineering
Technical University of Denmark
Kgs. Lyngby, Denmark

Co-supervisors

Prof. Torsten Dau
Dr. Sara Miay Kim Madsen
Hearing Systems Group
Department of Electrical Engineering
Technical University of Denmark
Kgs. Lyngby, Denmark

External advisor

Dr. Hamish Innes-Brown
The Bionics Institute of Australia
Melbourne, Australia

"Cuando se reflexiona sobre la curiosa propiedad que el hombre posee de cambiar y perfeccionar su actividad mental con relación a un objeto o problema profundamente meditado, no puede menos de sospecharse que el cerebro, merced a su plasticidad, evoluciona anatómica y dinámicamente, adaptándose progresivamente al tema. Esta adecuada y específica organización adquirida por las células nerviosas produce a la larga lo que yo llamaría talento profesional o de adaptación, y tiene por motor la propia voluntad, es decir, la resolución enérgica de adecuar nuestro entendimiento a la naturaleza del asunto."

"When one reflects on the ability that humans display for modifying and refining mental activity related to a problem under serious examination, it is difficult to avoid concluding that the brain is plastic and goes through a process of anatomical and functional differentiation, adapting itself progressively to the problem. The adequate and specific organization acquired by nerve cells eventually produces what I would refer to as professional or adaptational talent. As a motivator of the will itself, this brain organization provides the energy to adapt understanding to the nature of the problem under consideration."

Reglas y consejos sobre la investigacion biologica, 1898.
(Advice for a young investigator)
Santiago Ramón y Cajal (1852 - 1934).

Abstract

Cochlear implants (CIs) are a neural prosthesis that allows severely hearing-impaired listeners to achieve high levels of speech understanding in quiet environments. However, listening to a single person's voice among many, or to a melody in a complex musical arrangement, can be challenging for most CI listeners. In such scenarios, the listener needs to parse the sounds in the complex auditory scene, group them into meaningful auditory objects, or streams, and selectively attend to the stream of interest. Many studies have investigated the principles or cues that allow the healthy auditory system to perceptually group sounds into streams and selectively attend to one of them. However, the number of studies investigating these processes in CI listeners is limited and their findings are contradictory. The studies presented in this thesis aimed at improving our understanding of the perceptual organization of sounds by CI listeners. Behavioral and electrophysiological measures were combined to address the fundamental questions of whether CI listeners can perceptually segregate sequentially presented sounds and whether they can selectively attend to the stream of interest. The results showed that perceptual differences elicited by varying either the place or the pulse rate of the electrical stimulation allowed the listeners to perceptually group the sounds into auditory streams. The listeners were able to selectively attend to a target stream, and the effects of selective attention could be assessed using electroencephalography. However, the results also suggested that CI listeners might not be able to effectively suppress a competing stream and to initially select the stream of interest, which could contribute to the challenges experienced by CI listeners in complex listening scenarios. The findings from this thesis represent a valuable basis for future studies investigating the perceptual organization of sounds in CI listeners. In addition, these findings are relevant for the design of future devices, since they suggest that it may be possible to use brain signals to decode selective attention. This information could potentially be used to enhance the perceptual differences between the attended stream and the background sounds, aiding the listeners in complex auditory scenes.

Resumé

Et cochlear-implantat (CI) er et elektrisk apparat, der gør det muligt for mange svært hørehæmmede personer at opnå god taleforståelse i stille omgivelser. Dog kan det stadig være meget udfordrende for de fleste CI-brugere at høre en enkelt persons stemme blandt mange eller en melodi i et komplekst stykke musik. I sådanne scenarier skal lytteren analysere lydene i den komplekse auditoriske scene, gruppere dem i meningsfulde auditive objekter eller ”lydstrømme” og selektivt følge den lydstrøm, de er mest interesseret i. Mange studier har undersøgt de egenskaber, der giver det normale auditoriske system mulighed for perceptuelt at gruppere lyde i lydstrømme og selektivt rette opmærksomhed mod en af dem. Kun få studie har undersøgt disse processer i CI-brugere og deres resultater er modstridende. Formålet med undersøgelserne præsenteret i denne afhandling er at forbedre forståelsen af den perceptuelle organisering af lyd hos CI-brugere. Adfærdsmæssige og elektrofysiologiske målinger blev kombineret for svare på grundlæggende spørgsmål om, hvorvidt CI-brugere perceptuelt kan adskille lyde, der er præsenteret sekventielt, og om de selektivt kan fokusere deres opmærksomhed på de enkelte lydstrømme. Resultaterne viste, at perceptuelle forskelle fremkaldt ved at variere enten elektrodepositionen eller pulsfrekvensen af den elektriske stimulering tillod brugerne at gruppere lydene i auditoriske strømme. Brugerne kunne selektivt følge en bestemt lydstrøm, og effekten af selektiv opmærksomhed kunne måles ved hjælp af elektroencefalografi. Resultaterne demonstrerede dog også, at CI-brugeres evne til effektivt at kunne undertrykke en konkurrerende lydstrøm og fra starten rette deres opmærksomhed mod den lydstrøm, de var interesseret i muligvist er begrænset. Dette kunne bidrage til de udfordringer, som CI-brugere oplever i komplekse lydscenarier. Resultaterne fra denne afhandling er relevante for udformningen af fremtidige CI-apparater, da de antyder, at det kan være muligt at bruge hjernens signaler til at afkode selektiv opmærksomhed. Disse oplysninger kan potentielt bruges til at øge de perceptuelle forskelle mellem den ønskede lydstrøm og baggrundslyden, der hjælper brugere i komplekse auditoriske scenarier.

Acknowledgments

This thesis presents the result of three years of work, continuous learning and considerable personal development. This would not have been possible without the help, support and motivation I received from many people.

I would like to thank my supervisors Sara, Torsten and Jeremy, for their help and guidance. Sara, thank you for bearing with all my questions and complains, for always being the first one answering my emails and for the many discussions we had throughout the project. Torsten, thank you for taking me aboard the hearing systems group and for always keeping some time for me in your busy schedule. Jeremy, thank you for your patience, trust and understanding throughout the project, for giving me the freedom to chose the path while making sure that there was a way out.

I would also like to acknowledge the people who, not being officially a supervisor, have helped me in one or another way to make it through. Thank you, Golbarg, for your inspiration, courage and support, already well before I started this project. Thank you, Hamish, for the many discussions and the four incredible months at the Bionics Institute. Thank you, James, for your help and feedback. Thank you, Wiebke, for your support, generosity and for many helpful discussions. Thank you, Charlotte, for your help with the danish abstract. And thank you, Gerard, for always being there since the very first day I arrived to Denmark.

This work could not have been possible without the help from many volunteers who participated in the experiments here presented. Thank you all for your motivation and determination.

Big big thanks to my office mates and colleagues, both in Denmark and in Australia, for making this journey interesting and fun. Thanks to all the past and present members of the "microwave gang" for making sure I smiled and laughed every day.

I would also like to acknowledge the financial suport from the Oticon Centre of Excellence for Hearing and Speech Sciences (CHESS). Part of the research equipment for this project was provided by Cochlear Ltd. (Sydney, Australia). The head icon from the front page was designed by Smartline from flaticon.

Finally I would like to thank my parents, my little sister and Yuran for their unconditional love, trust, care and support.

Andreu Paredes Gallardo, May 28th, 2018.

Related publications

Journal papers

- Paredes-Gallardo, A., Madsen, S.M.K., Dau, T., Marozeau, J. (**2018a**). "The role of place cues in voluntary stream segregation for cochlear implant users", Trends Hear. 22, 1—13 https://doi.org/10.1177/2331216517750262
- Paredes-Gallardo, A., Madsen, S.M.K., Dau, T., Marozeau, J. (**2018b**). "The role of temporal cues in voluntary stream segregation for cochlear implant users", Trends Hear. 22, 1—13 https://doi.org/10.1177/2331216517750262
- Paredes-Gallardo, A., Innes-Brown, H., Madsen, S.M.K., Dau, T., Marozeau, J. (under review). "Electrode separation as a cue for auditory stream segregation in cochlear implant listeners: Evidence from behavioral measures and event-related potentials", Frontiers in Neuroscience - Auditory Cognitive Neuroscience

Conference papers

- Paredes-Gallardo, A., Madsen, S.M.K., Dau, T., Marozeau, J. (**2018c**). "The role of temporal cues on voluntary stream segregation in cochlear implant users", Proc. Int. Symp. Audit. Audiol. Res. Vol 6 Adapt. Process. Hear.

Published abstracts

- Paredes-Gallardo, A., Madsen, S.M.K., Dau, T., Marozeau, J. (**2016**). "Objective assessment of stream segregation abilities of cochlear implant users as a function of electrode separation", 9^{th} International Symposium on Objective Measures in Auditory Implants (OMAI), Szeged, Hungary, June 2016

- Paredes-Gallardo, A., Madsen, S.M.K., Dau, T., Marozeau, J. (**2016**). "Objective assessment of voluntary stream segregation abilities of cochlear implant users", 1^{st} Music and CI Symposium, Eriksholm Research Centre, Denmark, October 2016
- Paredes-Gallardo, A., Madsen, S.M.K., Dau, T., Marozeau, J. (**2016**). "Objective assessment of voluntary stream segregation abilities of cochlear implant users", Audiological Research Cores in Europe (ARCHES), 10^{th} meeting, Zurich, Switzerland, November 2016
- Paredes-Gallardo, A., Innes-Brown, H., Madsen, S.M.K., Dau, T., Marozeau, J. (**2017**). "The role of place and temporal cues on voluntary stream segregation of cochlear implant users", Conference on Implantable Auditory Prostheses (CIAP), Lake Tahoe, CA (USA), July 2017
- Paredes-Gallardo, A., Innes-Brown, H., Madsen, S.M.K., Dau, T., Marozeau, J. (**2018**). "Attentional modulation of event related potentials as a measure of stream segregation for cochlear implant listeners", Association for Research in Otolaryngology (ARO), 41^{st} Mid-Winter Meeting, San Diego, CA (USA), February 2018

Datasets

- Paredes-Gallardo, A., Madsen, S.M.K., Dau, T., Marozeau, J. (**2018**). "The role of place cues in voluntary stream segregation for cochlear implant listeners [Data set]." Zenodo. http://doi.org/10.5281/zenodo.890791
- Paredes-Gallardo, A., Madsen, S.M.K., Dau, T., Marozeau, J. (**2018**). "The role of temporal cues in voluntary stream segregation for cochlear implant listeners [Data set]." Zenodo. http://doi.org/10.5281/zenodo.1126666
- Paredes-Gallardo, A., Innes-Brown, H., Madsen, S.M.K., Dau, T., Marozeau, J. (**2018**). "Electrode separation as a cue for auditory stream segregation in cochlear implant listeners: Evidence from behavioral measures and event-related potentials [Data set]." Zenodo. http://doi.org/10.5281/zenodo.1211575

Additional journal papers

- Paredes-Gallardo, A., Epp, B., Dau, T. (**2016**). ”Can place-specific cochlear dispersion be represented by auditory steady-state responses?.” Hearing Research 335, 76−82 https://doi.org/10.1016/j.heares.2016.02.014
- Mehraei, G., Paredes-Gallardo, Shinn-Cunningham, B.G., Dau, T. (**2017**). ”Auditory brainstem response latency in forward masking, a marker of sensory deficits in listeners with normal hearing thresholds.” Hearing Research 346, 34−44 https://doi.org/10.1016/j.heares.2017.01.016

Contents

1

General introduction

"Since Aristotle, many philosophers and psychologists have believed that perception is the process of using the information provided by our senses to form mental representations of the world around us. [...] The job of perception, then, is to take the sensory input and to derive a useful representation of reality from it."

Bregman, 1990

Among the human senses, hearing provides us with a constant connection to the physical world. Unlike the visual system, the auditory system is constantly monitoring our surroundings. Hearing represents, moreover, an important aspect of human interaction and communication through speech and music. Thus, the consequences of hearing impairment go beyond the sensory deficit and can lead to a sense of profound social isolation (Wilson et al., 2017). In addition, it has been suggested that hearing loss can prevent or delay the acquisition of spoken language by children (e.g. Tomblin et al., 2015; Wilson et al., 2017) and several studies report a large gap between the academic achievements of hearing-impaired students and their normal-hearing (NH) peers (e.g. Qi and Mitchell, 2012). Therefore, many deaf or severely hearing-impaired people have benefited from cochlear implants (CIs), a sensory-neural prosthesis that can partially restore some auditory capacities.

1.1 The cochlear implant

The CI bypasses the outer, middle and inner ear, providing direct electrical stimulation close to the auditory nerve. The CI consists of two parts: The sound processor and the implant. The sound waves are captured by the microphones of the sound processor, where the acoustic signal is converted to a digital signal and encoded as a radio frequency signal. This signal is then transmitted through the skin to the implant, surgically placed between the skin and the skull. The implant decodes the radio frequency signal and converts it to electric current pulses, which are sent through thin wires to a small electrode array, surgically inserted in the cochlea (for a detailed review about the design of CIs, see Zeng et al., 2008). The current pulses, sent through the electrodes, stimulate the auditory nerve, resembling the transduction mechanism of the inner hair cells of a healthy cochlea, which transform the sound energy to spike trains.

The manipulation of either the stimulation electrode or the pulse rate is known to elicit perceptual differences in CI listeners (e.g. Eddington et al., 1978; Landsberger et al., 2016; Shannon, 1983). Thus, by using the envelope from different bands of the frequency spectrum to modulate the signal delivered by each electrode, the CI recreates an approximation of the frequency-to-place mapping (i.e. tonotopic organization) of the healthy cochlea (see Oxenham, 2018, for a review about the neural coding of sound). The stimulation pulse rate is kept constant in most of the current stimulation strategies

With a CI, an array of 12-24 electrodes is intended to replace the transduction from over 3000 inner hair cells. Albeit limited, the spectral resolution provided by current CIs may be enough to convey speech information (e.g. Shannon et al., 1995). Therefore, many CI users achieve high levels of speech understanding in quiet (e.g. Zeng et al., 2008). However, CI users typically experience difficulties to understand a single person's voice among many, or to recognize a familiar melody in a complex musical arrangement (e.g. Gfeller et al., 2005; Nelson et al., 2003).

1.2 Auditory scene analysis

In many natural listening scenarios, the sounds from multiple sources compose a complex acoustic waveform. Thus, the information about the individual sources is not explicitly present in the signal that reaches the ears. Nevertheless, NH listeners are generally able to hear an individual sound in this mixture. The task of selectively listen to a specific source in the presence of others is often referred to as *the cocktail party problem* (Cherry, 1953). In such scenario, the first challenge for the listeners is to perceptually organize the mixture of sounds into auditory objects or *streams*, a process known as *auditory scene analysis* (Bregman, 1990). The second challenge of the cocktail party problem is to selectively attend to the stream of interest and ignore the others (McDermott, 2009). However, the two problems are closely related since, as will be discussed further below, selective attention can influence the perceptual organization of the sounds (e.g. Van Noorden, 1975; see also the temporal coherence model proposed by Shamma et al.,2011).

Auditory scene analysis includes the perceptual organization of both simultaneously presented sounds (e.g. Micheyl and Oxenham, 2010b) and sequentially presented sounds (e.g. Moore and Gockel, 2012). The role of simultaneous grouping is to determine which frequency components should be allocated to each auditory stream at a given time instant. Thus, the precise and accurate spectral decomposition of the incoming signal is necessary for simultaneous grouping. The role of sequential grouping is, instead, to determine which sounds should be allocated to each auditory stream over time. It is apparent that both simultaneous and sequential grouping interact with each other, however, they are often investigated separately.

A mixture of sounds can be decomposed in many different ways. Therefore, the auditory system makes use of some principles, or grouping *cues*, to solve the problem of scene analysis. For example, sound events originated by the same source generally start and stop synchronously and exhibit coherent amplitude fluctuations. Consequently, synchronous sounds or coherently fluctuating sounds are generally grouped into a single stream (i.e. stream *integration*). Conversely, asynchronous sounds or incoherently fluctuating sounds are likely to be grouped into multiple streams (i.e. stream *segregation*) (e.g. Bregman, 1990, chapter 3).

The perceptual organization of sequential sounds depends, to some extent, on the degree of perceptual continuity between consecutive sounds, equivalent to the *good continuation* principle proposed by the Gestalt psychologists (Koffka, 1935). Thus, similar and/or slowly-varying sound events are likely to be integrated into a single auditory stream whereas different and/or fast-varying sounds are likely to be segregated (e.g. Bregman, 1990, chapter 2). In NH listeners, sequential stream segregation has been investigated on the basis of acoustic properties, such as the frequency content (e.g. Bregman and Campbell, 1971; Van Noorden, 1975), the temporal envelope (e.g. Cusack and Roberts, 2000; Iverson, 1995; Vliegen et al., 1999), the spatial characteristics (e.g. David et al., 2015; Sach and Bailey, 2004; Stainsby et al., 2011) or the intensity of the sounds (e.g. Van Noorden, 1977; Van Noorden, 1975). Overall, large differences between consecutive sound events lead to abrupt perceptual discontinuities and therefore promote stream segregation. In contrast, small differences between consecutive sounds lead to a more continuous percept and therefore, promote stream integration. Some studies have also suggested that temporal coherence can be a strong grouping cue for sequentially presented sounds (e.g. Christiansen et al., 2014).

Attention plays an important role in auditory scene analysis, and it can bias the perceptual organization of sounds. This was shown by Van Noorden (1975), who measured the segregation threshold for sequences of alternating low- and high-frequency pure tones that repeated over time. He observed that a smaller frequency difference was needed when the listeners were instructed to segregate the sounds than when they were instructed to integrate them. Consequently, Van Noorden defined two perceptual boundaries: the *fission boundary* (FB), representing the frequency separation at which the tones could no longer be segregated, and the *temporal coherence boundary* (TCB), representing the frequency separation at which the tones could no longer be integrated. Whereas the FB showed little dependency on the presentation rate of the tones, the TCB increased for faster tone presentation rates. The dependency of the TCB on the tone presentation rate was consistent with the good continuation principle from the Gestalt psychologists, since fast alternating sounds were more likely to result in a less continuous percept than slowly-varying sounds. For frequency separations above the FB and below the TCB, the sounds could either be integrated or segregated, which can result in multiple perceptual switches between an integrated and a segregated percept, often referred to as a *bistabile* percept (e.g. Pressnitzer and Hupé, 2006). Furthermore, the perceptual organization of the tones in this ambiguous region could be biased by the intention of the listener. The role of attention in auditory scene analysis has been highlighted by the temporal coherence model (Shamma et al., 2011; Shamma et al., 2013), which suggests that the process of object formation is initiated by attention: when attention is focused on a given auditory feature all coherent perceptual features are bond together into a stream.

The duration of the auditory stimulus can also affect its perception. Specifically, the probability of perceptually segregating a sequence of sound events generally increases over time. This phenomenon is often referred to as the build-up effect (e.g. Anstis and Saida, 1985; Bregman, 1990; Moore and Gockel, 2012). The time required for stream segregation to occur depends on the exact stimulus parameters or the specific paradigm, but is typically in the order of several seconds (e.g. Anstis and Saida, 1985; Bregman, 1978).

1.3 Assessment of auditory stream segregation

In humans, auditory stream segregation has been assessed using both behavioral and non-invasive neurophysiological methods. Behavioral assessments are often carried out by asking the listener to report whether a particular sound sequence was integrated or segregated. In this subjective approach, the listener typically undergoes some training to distinguish the one-stream and the two-stream percepts. An alternative approach has been to measure the performance of the listener in a given task (e.g. a signal detection or discrimination task) that is affected by the integration or segregation of the sounds. For example, the listeners are better at making temporal judgements between sounds when they are grouped into a single stream than when they are grouped into multiple streams (e.g. Bregman and Campbell, 1971; Micheyl and Oxenham, 2010a; Roberts et al., 2002, 2008; Van Noorden, 1975; Vliegen et al., 1999). Since this approach does not rely on subjective reports of perceived segregation it has been referred to as a behavioral and "objective" measure of auditory stream integration/segregation (Micheyl and Oxenham, 2010a).

Within the behavioral and objective measures of auditory stream segregation, a distinction can be made between those in which the performance in the task is facilitated by the segregation of the sounds and those in which it is worsened by segregation. In the first case, the attention of the listeners is biased towards a segregated percept which, in the context of this thesis, will be referred to as *voluntary stream segregation*. In the latter case, the attention of the listeners is biased towards integration, here referred to as *obligatory stream segregation*. Thus, the FB and the TCB represent the thresholds of voluntary and obligatory stream segregation, respectively.

Auditory stream segregation has also been investigated using a variety of non-invasive neurophysiological methods, such as electroencephalography (EEG) or magnetoencephalography (MEG) (for a review, see Alain et al., 2015; Gutschalk and Dykstra, 2014; Snyder and Alain, 2007). Several EEG studies have recorded the mismatch negativity (MMN) or the P300 response as indexes of stream segregation (e.g. Nie et al., 2014; Sussman et al., 1999). Both the MMN and the P300 reflect a change detection process (e.g. Näätänen et al., 2007; Sutton et al., 1965), and they are generally recorded using an oddball paradigm. In the context of stream segregation, the stimuli used to elicit the MMN or the P300 response generally consist of a sequence of sounds where a deviant can only be detected if the sounds are segregated in multiple streams. Conversely, the resulting pattern of an integrated percept may be too complex for the deviant to be detected (e.g. Nie et al., 2014; Snyder and Alain, 2007; Sussman et al., 1999). Both the MMN and the P300 responses reflect the processing of deviant patterns and therefore, their relation to stream segregation is indirect (i.e. they do not monitor the ongoing stream/s). Another approach used by MEG and EEG studies considers the relation between the recorded long-latency neural responses (i.e. > 60 ms) and the corresponding subjective reports of perception from the listeners (e.g. Gutschalk, 2005; Hill et al., 2012; Snyder et al., 2006). In these studies, the stimuli consist of simple sequences of repeating A and B tone triplets (ABA). The results typically show an enhanced neural response to the B tones when the sounds are perceptually segregated, suggesting a strong relation between the cortical responses and the perceptual organization of the sounds.

In recent years, several studies have used recordings of event-related responses (ERPs) to investigate the process of object selection and selective attention, tightly related to stream segregation (e.g. Choi et al., 2013, 2014; Dai and Shinn-Cunningham, 2016; Dai et al., 2018). Auditory selective attention enhances the amplitude of the N1 component from the ERPs in response to the attended sounds and suppresses those to the ignored sounds (e.g. Hillyard et al., 1998; Hillyard et al., 1973; Picton and Hillyard, 1974). Unlike paradigms making use of change detection components such as the MMN or the P300, measures of the N1 attentional modulation monitor the ongoing streams, providing insights about the time course of auditory attention. Moreover, the attentional modulation of the N1 response can be recorded using a wide variety of paradigms.

1.4 Auditory stream segregation for cochlear implant users

Previous studies have investigated both simultaneous and sequential stream segregation in CI listeners. Carlyon et al. (2007) and Cooper and Roberts (2010) assessed the effect of across-channel pulse rate differences, sound onset asynchrony and abrupt sound intensity changes in simultaneous stream segregation for CI listeners. They found no effect of pulse rate differences, even though the listeners could use large onset asynchronies or large changes in level to segregate concurrent sounds. However, the effects of these cues were relatively small, suggesting that CI listeners might not be able to benefit from them in complex listening scenarios.

Several studies have investigated the role of different cues in sequential stream segregation for CI listeners. Their findings have been contradictory: whereas some suggest that CI listeners can perceptually segregate sequential sounds (Böckmann-Barthel et al., 2014; Chatterjee et al., 2006; Duran et al., 2012; Hong and Turner, 2006, 2009; Innes-Brown et al., 2011; Marozeau et al., 2013; Tejani et al., 2017), other studies challenge these findings (Cooper and Roberts, 2007, 2009). Such disagreements could be partially related to methodological differences between the studies, which makes the comparison of the results difficult (e.g. Nie and Nelson, 2015). A summary of the methods from the previous studies is given in table 1.1. For example, while some studies assessed stream segregation using subjective reports of perception from the listeners, others used objective measures. Some studies assessed obligatory stream segregation, some assessed voluntary stream segregation and others did not provide specific listening instructions. Also, some studies presented acoustic stimuli to the listeners while others bypassed the listener's sound processor, directly stimulating the implant. Moreover, phenomena such as the build-up effect, considered *"an essential landmark of auditory stream segregation"* (Hupé and Pressnitzer, 2012), remain poorly understood in CI listeners (Böckmann-Barthel et al., 2014; Chatterjee et al., 2006; Cooper and Roberts, 2009).

Table 1.1: Summary of the previous studies investigating stream segregation in CI listeners.

Study	Direct reports of perception	Obligatory / Voluntary	Acoustic stimuli
Chatterjee et al. (2006)	Yes	No instructions	No
Hong and Turner (2006)	No	Obligatory	Yes
Cooper and Roberts (2007)	Yes	No instructions	No
Cooper and Roberts (2009)	No	Both	No
Hong and Turner (2009)	No	Voluntary	Yes
Innes-Brown et al. (2011)	Yes	Voluntary	Yes
Duran et al. (2012)	No	Obligatory	No
Marozeau et al. (2013)	Yes	Voluntary	Yes
Chatterjee et al. (2006)	Yes	No instructions	No
Böckmann-Barthel et al. (2014)	Yes	No instructions	Yes
Tejani et al. (2017)	No	Obligatory	No

The investigation of the perceptual organization of sounds in CI listeners has relied on behavioral measures, and no attempts have previously been made to assess auditory stream segregation and selective attention using neurophysiological methods. To some extent, this could be due to the limitations imposed by the presence of a magnet in the implant and the strong electrical signals resulting from the radio frequency communication between the sound processor and the implant.

1.5 Aims of the thesis

The work presented here aimed at improving the understanding of the perceptual organization of sound by CI listeners. The research projects described throughout the chapters of this thesis address the fundamental questions of whether CI listeners can segregate sequential sounds, whether a two stream percept builds up over time and whether they are able to selectively attend to the object of interest. Thus, voluntary stream segregation and the FB (i.e. the smallest difference between the sounds which allows their segregation) are investigated in this thesis.

It has been suggested that the discrepancies between the findings from the previous studies may have arisen, at least partially, from the uncertainty linked to the use of subjective reports of perception with CI listeners (e.g. Chatterjee et al., 2006; Cooper and Roberts, 2007; Hong and Turner, 2009) or from the lack of control over the signal delivered to the listeners when the sounds are presented via the listener's speech processor (e.g. Cooper and Roberts, 2009; Tejani et al., 2017). Therefore, the experimental paradigms described in the projects of this thesis are designed to investigate stream segregation and selective attention through behavioral, yet objective, measures, where perceptual differences are elicited via direct stimulation of the CI, bypassing the listener's speech processor. Moreover, the possibility of assessing auditory stream segregation and selective attention in CI listeners using EEG recordings is investigated.

1.6 Overview of the thesis

Chapter 2 describes the role of place cues in voluntary stream segregation for CI listeners using sequences of alternating and repeating sounds. Perceptual differences between the sounds are elicited by stimulating different electrodes at a constant pulse rate. Stream segregation is assessed indirectly using a temporal delay detection task, avoiding the uncertainty of direct reports of perception. A first experiment is performed to clarify whether CI listeners can use electrode separation as a cue to segregate the two sounds, and whether a segregated percept builds up over time. A second experiment is performed on a subset of the listeners to estimate the FB on the basis of electrode separation.

Chapter 3 describes the role of temporal cues in voluntary stream segregation for CI listeners using the delay detection task described in chapter 2. Instead of varying the stimulation electrode, perceptual differences between the sounds are here elicited by varying the pulse rate at a fixed cochlear location. The performance in the delay detection task is used to clarify whether CI listeners can use pulse rate differences as a cue to segregate the sounds and whether a two-stream percept builds up over time. The FB is estimated on the basis of pulse rate differences. Furthermore, the contribution of place vs temporal cues for voluntary stream segregation is assessed by combining the results from chapters 2 and 3 on a common perceptual scale.

Chapter 4 describes the role of place cues in voluntary stream segregation for CI listeners when both the target and the distractor streams are composed of multiple and different sounds. A behavioral deviant detection task is used to assess whether CI listeners can use electrode separation as a cue to segregate sounds in a more complex scenario than the one considered in chapters 2 and 3. Furthermore, the behavioral deviant detection task is combined with ERP recordings to assess whether selective auditory attention modulates the amplitude of the N1 response in CI listeners. If this would be the case, it could provide a better insight about the process of selective attention and object selection. Finally, chapter 4 aims to determine whether the N1 attentional modulation can be used as an objective tool to assess auditory stream segregation in CI listeners.

Finally, the main findings of each chapter are summarized, and their possible implications are discussed in chapter 5.

2

The role of place cues in voluntary stream segregation[a]

Abstract

Sequential stream segregation by cochlear implant (CI) listeners was investigated using a temporal delay detection task composed of a sequence of regularly presented bursts of pulses on a single electrode ("B") interleaved with an irregular sequence ("A") presented on a different electrode. In half of the trials, a delay was added to the last burst of the regular B sequence and the listeners were asked to detect this delay. As a jitter was added to the period between consecutive A bursts, time judgments between the A and B sequences provided an unreliable cue to perform the task. Thus, the segregation of the A and B sequences should improve performance. In experiment 1, the electrode separation and the sequence duration were varied to clarify whether place cues help CI listeners to voluntarily segregate sounds and whether a two-stream percept needs time to build up. Results suggested that place cues can facilitate the segregation of sequential sounds if enough time is provided to build up a two-stream percept. In experiment 2, the duration of the sequence was fixed and only the electrode separation was varied in order to estimate the fission boundary. Most listeners were able to segregate the sounds for separations of three or more electrodes and some listeners could segregate sounds coming from adjacent electrodes.

[a] This chapter is based on Paredes-Gallardo et al. (2018a).

2.1 Introduction

Current cochlear implant (CI) stimulation strategies convey spectral information through place cues, with different frequency bands stimulating different electrodes (e.g. Zeng et al., 2008). However, it is not known to what extent CI listeners can make use of electrode separation cues to segregate sounds. Findings from previous studies have been contradictory. Some studies found similar trends as in notmal-hearing (NH) listeners (Böckmann-Barthel et al., 2014; Chatterjee et al., 2006; Hong and Turner, 2006; Tejani et al., 2017) whereas other studies did not find any effect of the sequence duration or the tone presentation rate (Cooper and Roberts, 2007, 2009) which are well documented in studies with NH listeners (e.g. Bregman, 1990; Moore and Gockel, 2012; Van Noorden, 1975). Thus, it has been suggested that CI listeners might only experience some aspects of stream segregation as a function of electrode separation (Chatterjee et al., 2006; Cooper and Roberts, 2007, 2009; Hong and Turner, 2006; Tejani et al., 2017).

Previous behavioral studies assessing auditory stream segregation abilities of CI listeners as a function of place cues have made use of both subjective (Böckmann-Barthel et al., 2014; Chatterjee et al., 2006; Cooper and Roberts, 2007) and objective measures (Cooper and Roberts, 2009; Hong and Turner, 2006; Tejani et al., 2017). The subjective measures require the listener to be able to experience both fused and segregated percepts. It is unclear whether CI listeners can experience both fused and segregated percepts during their training sessions and thus, results from the subjective measures could reflect electrode discrimination instead of perceived segregation (Cooper and Roberts, 2007).

Most of the objective studies assessing stream segregation abilities of CI listeners used the irregular rhythm detection task (Cusack and Roberts, 2000; Roberts et al., 2002). In this task, listeners are presented with sequences of alternating A and B tones. In some of the sequences, the timing between A and B sounds is kept constant throughout, while in other sequences, the B tones are gradually delayed along the sequence. Listeners are asked to decide if a given sequence has an irregular rhythm. Since the detection of rhythm changes is more difficult when the A and B sounds fall in separate streams (e.g. Micheyl and Oxenham, 2010b; Van Noorden, 1975), the integration of the streams improves the performance in the detection task. Studies using the irregular rhythm detection task with CI listeners (Cooper and Roberts, 2009; Hong and Turner, 2006; Tejani et al., 2017) observed better performance for small rather than large electrode separations. However, the results also presented substantial nonmonotonicities (Tejani et al., 2017), and a build-up effect of streaming was not found (Cooper and Roberts, 2009). The irregular rhythm detection task has one confounding factor: several studies have suggested that temporal gap detection abilities in CI listeners worsen when the gap markers are presented from different electrodes (e.g Hanekom and Shannon, 1998; Wieringen and Wouters, 1999) or with different pulse rates (Chatterjee et al., 1998). Thus, a worsening of the detection performance on the irregular rhythm detection task might not be solely due to stream segregation (Cooper and Roberts, 2009; Hong and Turner, 2006; Tejani et al., 2017).

While the irregular rhythm detection task has been used to assess obligatory stream integration abilities and the temporal coherence boundary, voluntary stream segregation has received less attention. In one experiment, Cooper and Roberts (2009) assessed the effect of electrode separation on the ability to segregate a simple melody from interleaved distractor notes. The task was facilitated by the segregation of the streams and, thus, assessed voluntary segregation. They observed that CI listeners were not able to identify the target melody in the presence of the interleaved distractors without loudness cues, regardless of the electrode range of the distractors relative to the melody. The sequences used by Cooper and Roberts (2009) had a fixed duration of 2.2 seconds. It is therefore unclear whether the poor performance in the task was due to poor voluntary stream segregation abilities or due to too short sequences, assuming that CI listeners might need more time to build-up a two stream percept even in a segregation-promoting paradigm.

The present study investigated voluntary stream segregation abilities in CI listeners as a function of place cues. Rhythm-detection performance was measured in a paradigm where the listeners were required to make within-stream time judgements in the presence of a temporally irregular distractor stream. Thus, the task became easier if the listeners could segregate the target from the distractor. This paradigm has previously been used with NH listeners (Micheyl and Oxenham, 2010b; Nie and Nelson, 2015; Nie et al., 2014) but not yet considered in studies with CI listeners. While in the irregular rhythm detection task (Cusack and Roberts, 2000) the integration of the streams improves performance, in the present study, the segregation of the streams should facilitate detection performance. Thus, the gap detection confounding factor of the irregular rhythm detection task is here avoided by encouraging the listeners to perform within-channel temporal judgements. In experiment 1, the electrode separation and the sequence duration were varied to clarify 1) whether place cues help CI listeners to voluntarily segregate sounds and 2) whether a two-stream percept needs some time to build up. Experiment 2 combined measurements at three extra electrode separations in a subset of the listeners with an ideal observer (IO) model to estimate the minimum electrode separation needed to segregate the streams.

2.2 Experiment 1: Exploring the contribution of place cues to voluntary stream segregation

2.2.1 Rationale

Experiment 1 aimed to determine whether place cues can help CI listeners to voluntarily segregate sequential sounds and whether this segregation occurs instantaneously or if it needs some time to build up. Stream segregation abilities of CI listeners were assessed in a rhythm detection task. The paradigm was inspired by Micheyl et al. (2005) and Micheyl and Oxenham (2010b), and has previously been used by Nie et al. (2014) and Nie and Nelson (2015) to assess voluntary stream segregation abilities of NH listeners. In this paradigm, the listeners are asked to detect a small delay applied to the last sound of the sequence. The rhythm detection task is facilitated by the segregation of the streams. Thus, if place cues help CI listeners to segregate the A and B sounds, better performance should be achieved for larger electrode separations between the sounds. Conversely, if place cues do not contribute to the segregation of the sounds, the performance in the rhythm detection task should not depend on the electrode separation between the streams. Furthermore, the presence of a build-up effect should result in better performance for the longer sequences, whereas the lack of such build-up should lead to similar performance for short and long sequences. The better performance for the longer sequences could also reflect the fact that the listeners have longer time to focus on the steady rhythm of the target stream. Thus, rhythm detection performance was also measured for the long and short sequences in the absence of the distractor stream, to quantify the effect of sequence duration on the task when no stream segregation was necessary.

2.2.2 Methods

Listeners

Nine Cochlear CI listeners (six female and three male) participated in this experiment. The listeners were aged between 19 and 78 years (mean: 48 years, SD: 25 years; see table 2.1.) and had no residual hearing in their implanted ear. All listeners were bilateral except listener 7 who was bimodal. For listener 7, the contralateral ear was unaided and blocked with an ear plug during the experiments. All listeners provided informed consent prior to the study and all experiments were approved by the Science-Ethics Committee for the Capital Region of Denmark (reference H-16036391).

Table 2.1: Relevant information about the nine CI listeners. *M* stands for male and *F* for female.

ID	Age	Gender	Onset of deafness	Implant (ear)	Years of experience	Exp 1	Exp 2
CI 1	19	F	Prelingual	CI24RE (right)	16	Yes	No
CI 2	21	F	Prelingual	CI24R (right)	14	Yes	No
CI 3	21	M	Prelingual	CI24RE (right)	9	Yes	Yes
CI 4	74	F	Postlingual	CI24R (left)	13	Yes	Yes
CI 5	73	M	Postlingual	CI24RE (right)	3	Yes	Yes
CI 6	64	F	Perilingual	CI24R (right)	15	Yes	Yes
CI 7	78	M	Postlingual	CI24RE (right)	3	Yes	No
CI 8	61	F	Perilingual	CI24RE (right)	3	Yes	Yes
CI 9	21	F	Prelingual	CI24RE (left)	16	Yes	Yes

Stimuli and conditions

The stimulation paradigm is illustrated in figure 2.1, where different panels represent different conditions. A sequence of regularly presented bursts of pulses on a single electrode (B) was interleaved with an irregular sequence presented on a different electrode (A). In half of the trials, a small temporal delay (Δt) was added to the last burst of the regular B sequence, the target stream. The listeners were asked to indicate after each trial whether or not the last sound of the sequence was delayed. A jitter was added to the period between consecutive bursts of the A sequence (i.e. the distractor stream) making time judgments between successive A and B sounds an unreliable cue for performing the task. Therefore, to optimize performance, the listener needed to compare the time interval between the last two B-sounds with those between previous B-sounds. Thus, the task became easier when the A and B sequences felt into different streams (Micheyl and Oxenham, 2010b; Nie and Nelson, 2015; Nie et al., 2014), encouraging the listener to segregate the sounds.

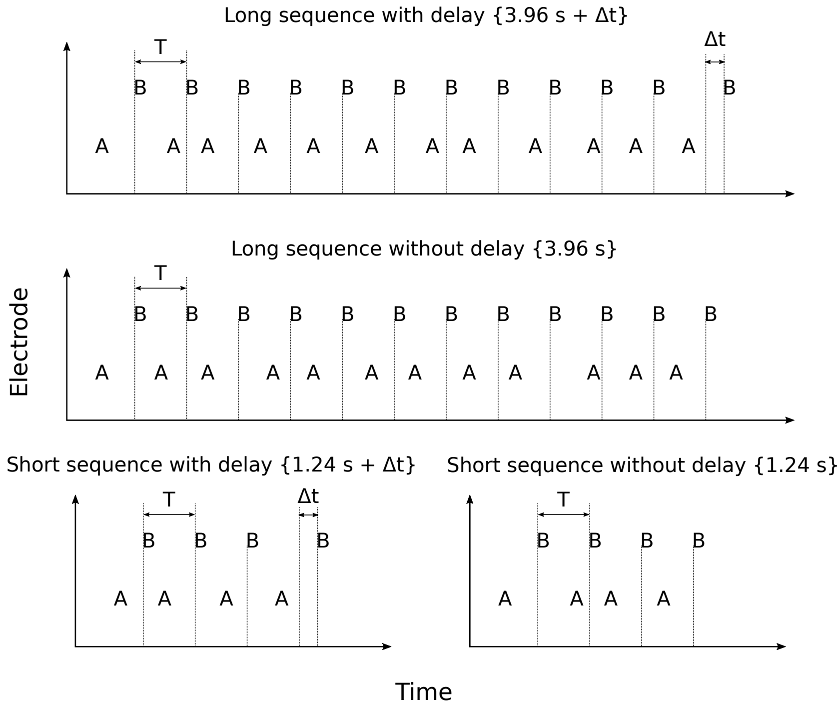


Figure 2.1: Graphical representation of the experimental paradigm. The onset-to-onset interval is represented by T and the delay of the last B sound by Δt. The electrode separation between A and B sounds varied across conditions.

Two sequence durations were tested (figure 2.1). The long sequence consisted of 12 AB pairs and the short sequence of 4 AB pairs, resulting in a nominal duration of 3.96 and 1.24 s, respectively, when no Δt was present. All sequences started with the distractor stream (A). The target stream (B) was always played through electrode 11[b], located at the mid-point of the array, with an onset-to-onset interval of 340 ms. The distractor stream (A) was played through either electrode 12 or 19 depending on the condition, leading to an electrode sep-

[b] In the Cochlear electrode array, electrode 1 is the most basal electrode and electrode 22 the most apical one.

aration between target and distractor of either one or eight electrodes in the apical direction. This choice aimed to make the listening task more pleasant for the listeners by avoiding basal, high-pitch electrodes. The onset-to-onset interval of the distractor stream varied for each presentation, having a nominal duration of 340 ms ± 220 ms jitter. The jitter values were uniformly distributed. Consecutive A and B sounds were always separated by a minimum interval of 10 ms.

Each A and B sound consisted of a 50-ms biphasic pulse burst presented with a fixed rate of 900 pulses per second (pps) in monopolar mode. Each biphasic pulse had a phase width of 25 µs and phase gap of 8 µs. The stimuli were presented through the Nucleus Implant Communicator research interface (NIC v2, Cochlear Limited, Sydney).

Rhythm detection performance for the long and short sequences was also measured without the distractor stream. These conditions were significantly easier than the test conditions and thus, a different (shorter) Δt value was used to avoid ceiling effects. Since listener 2 was not available for the control condition, no control data were available for this listener.

For each combination of electrode separation and sequence duration, 60 presentations of the delayed sequence and 60 presentations of the non-delayed sequence were used to calculate the listener's sensitivity (d') to the delayed target.

Loudness balancing

Loudness has been found to be an effective cue for sound segregation of CI listeners (e.g. Cooper and Roberts, 2009; Marozeau et al., 2013). The stimuli were therefore loudness-balanced in this experiment. Categorical loudness scaling was performed for each electrode using an 11-step attribute scale ranging from "off" (attribute 0) to "too loud" (attribute 10). The intensity of the pulse train was increased in steps of 1.6 dB until the listener could perceive a "just noticeable" sound (attribute 1). The intensity of the pulse train was further increased with a step size of 0.8 dB until the sound became "comfortable but soft" (attribute 5). Finally, a step size of 0.3 dB was used until the sound became "loud but comfortable" (attribute 7) and then decreased again until the "most comfortable" level was reached (MCL, attribute 6).

Once all electrodes were set at MCL, each pair of target and distractor electrodes (i.e. 11/12 and 11/19) were loudness matched by the listener using a simple user interface which allowed the increase and decrease of the distractor sound intensity in steps of 0.15, 0.3 or 0.45 dB. The loudness matching of the electrode pairs was performed in the beginning of each session. The level of the loudness balanced stimuli did not markedly change for the different sessions.

Delay adjustment procedure

Individual Δt values were used in this study. Δt values were chosen such that all listeners would be equally sensitive to the delayed target in a given condition. The long sequence with the largest electrode separation (12 AB pairs with the distractor stream played at electrode 19) was used for the individual adjustment of Δt. The sensitivity to the delayed target was measured for four different delays: 5, 40, 80 and 120 ms or 5, 30, 60, 90 ms (listener 9) based on 30 presentations of each delayed sequence and 30 presentations of the non-delayed sequences. The four Δt values were presented in random order. A sigmoid function bounded between 0 and 4.7 was fitted to the data of each listener using the Matlab fitting toolbox. The individual Δt was defined as the delay leading to a signal sensitivity of d' = 2. Individual Δt values were always smaller than the 110 ms jitter applied to each A sound (see table 2.2).

The same delay adjustment procedure was used to find the individual Δt values to be used in the control conditions. In this case, the long sequence without the distractor sounds and with delays of 5, 20, 40 and 60 ms was used to fit the psychometric function. The delay leading to d' = 3 was chosen as Δt for the control condition (see table 2.2). This d' value was chosen to keep the control conditions relatively easy while avoiding ceiling effects.

Procedure

The experiments took place in a double-walled, sound-attenuating booth at the Technical University of Denmark, and were organized in two sessions, each lasting 2h including short breaks. The first session included a brief description of the task, the loudness balancing, training for the rhythm detection task and the delay adjustment procedures. All four conditions as well as the two control conditions were tested in the second session.

A one-interval, two-alternative, forced-choice procedure was used, where the listeners were asked to report whether a given sequence contained a delayed target or not. A one-interval task was chosen instead of a two- or three-interval

Table 2.2: Individual Δt values as obtained from the delay adjustment procedure.

ID	Δt [ms] for $d' = 2$	Δt [ms] for control condition, $d' = 3$
CI 1	40	30
CI 2	70	-
CI 3	52	35
CI 4	45	35
CI 5	35	32
CI 6	80	55
CI 7	80	80
CI 8	60	28
CI 9	35	30

paradigm to minimize the attentional effort required to perform the task (Nie and Nelson, 2015).

Listeners were familiarized with the rhythm detection task by listening to the target stream in the absence of any distractor sound. They were asked to report whether the sequence of target sounds was regular (non-delayed) or irregular (delayed). Once the task was clear, the distractor stream was introduced from electrode 19 (i.e. a large electrode separation) at a soft (but audible) level. Listeners were asked to perform the task while ignoring the distractor sounds. The level of the distractor stream increased progressively until both target and distractor sounds were presented at the listener's MCL. The training procedure was repeated with the distractor presented at electrode 12 (i.e. a small electrode separation). The duration of the training varied across listeners, ranging between 10 to 20 min.

A total of eight different sequences were presented to the listeners, resulting from the combination of 2 possible distractor electrodes (12 or 19), two sequence durations (4 and 12 AB pairs) and two different Δt values (delayed or non-delayed). Short and long sequences were presented in different blocks. In each block, each of the four possible sequences was repeated 12 times in pseudo-random order, ensuring that the distractor electrode alternated from one sequence to the next one. Thus, the first sound of each sequence alternated between electrode 12 and 19, contributing to the resetting of the build-up of a two-stream percept after each presentation (Roberts et al., 2008). Each block was repeated five times in a random order.

The control conditions were tested in four blocks (two with long sequences and two with short sequences) containing 30 repetitions of the delayed and 30 repetitions of the non-delayed sequences. The control blocks were randomly presented at the beginning or at the end of either session.

Statistical analysis

Unless otherwise specified, statistical inference was performed by fitting a mixed-effects linear model to the computed d' scores. The experimental factors (i.e. electrode separation, sequence duration and their interaction) were treated as fixed effects terms whereas listener-related effects were treated as random effects. The model was implemented in `R` using the `lme4` library (Bates et al., 2014) and the model selection was carried out with the `lmerTest` library (Kuznetsova et al., 2017) following the backwards selection approach based on step-wise deletion of model terms with high p-values (Kuznetsova et al., 2015). The p-values for the fixed effects were calculated from F-tests based on Sattethwaite's approximation of denominator degrees of freedom and the p-values for the random effects were calculated based on likelihood ratio tests (Kuznetsova et al., 2015). Post-hoc analysis was performed through contrasts of least-square means using the `lsmeans` library (Lenth, 2016) and the `lme4` model object. P-values were corrected for multiple comparisons using the Tukey method.

2.2.3 Results

The individual results from experiment 1 are shown in figure 2.2, where each panel represents one listener. The sensitivity to the delayed B sound is plotted for each electrode separation and for the control condition with black circles representing the long sequence and gray triangles representing the short sequence.

Figure 2.3 shows the results from experiment 1. Panel A contains d' scores for all combinations of sequence duration and distractor electrode. Panel B shows the individual difference between d' scores in the long and short sequences, for each distractor electrode. Panel C shows the individual difference between d' scores obtained when the distractor and the target were separated by one and eight electrodes, for each sequence duration. The significance of the statistical contrasts is illustrated with asterisks. Both sequence duration [$F(1, 7.94) =$

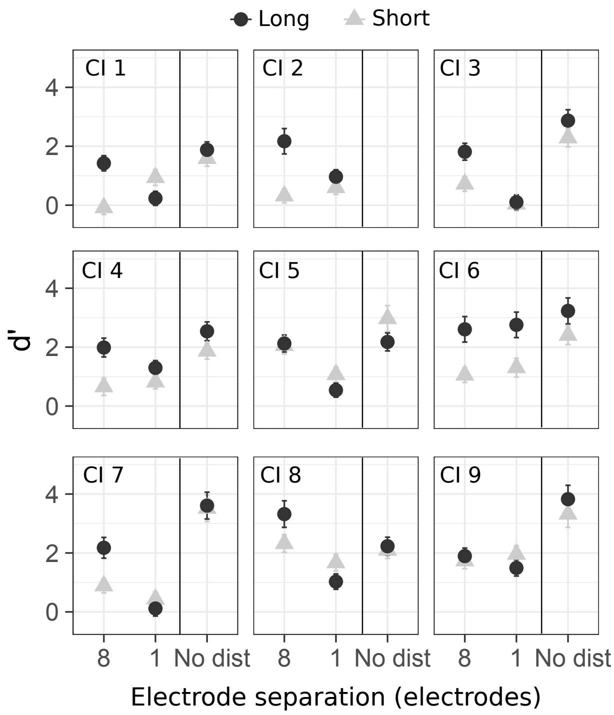


Figure 2.2: Individual sensitivity to the delayed tone (d') for each electrode separation and sequence duration. Error bars represent the standard errors of the d' estimates.

7.214, p = 0.028], distractor electrode [F(2, 7.85) = 16.348, p = 0.002] and their interaction [F(2, 15.18) = 17.503, p < 0.001] were found to be significant factors in the statistical model.

Panels A and C in figure 2.3 show that for the long sequence, greater d' scores were obtained when the electrode separation between distractor and target was eight electrodes rather than one [t(19.23) = 4.439, p = 0.003, difference estimate = 1.221], implying that CI listeners benefitted from the larger target-distractor electrode separation to perform the task. Conversely, the distractor electrode did not significantly affect d' scores in the short sequence [t(19.23) = 0.333, p = 0.999, difference estimate = 0.091].

Panels A and B in figure 2.3 show a significant difference in d' scores between the long and short sequences when distractor and target streams were separated by eight electrodes [t(14.49) = 5.311, p = 0.001, difference estimate = 1.096]. No significant difference was observed when distractor and target streams were separated by one electrode [t(14.49) = -0.160, p = 1.000, difference estimate = -0.033] or for the control condition [t(15.79) = 1.588, p = 0.533, difference estimate = 0.341].

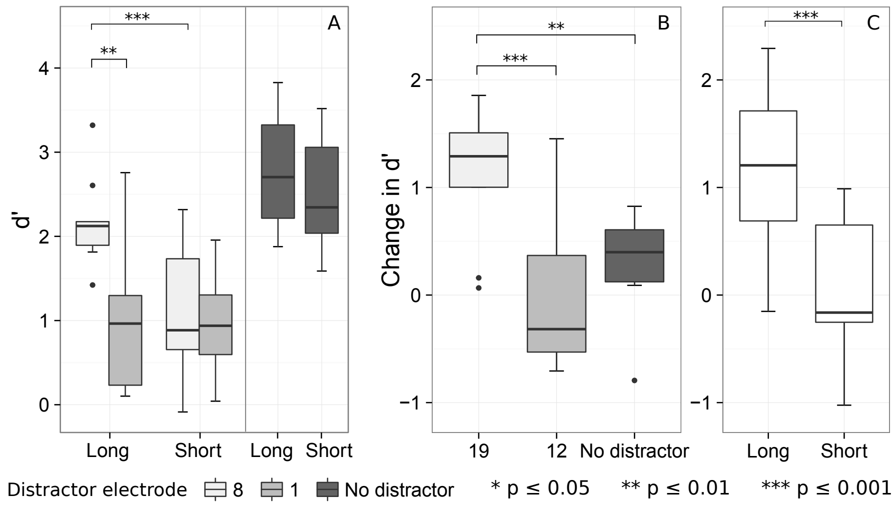


Figure 2.3: A] Sensitivity to the delayed tone (d') for each condition. B] Individual differences in d' between the long and short sequences for the different electrode separations and the control condition. C] Individual differences in d' achieved when the distractor and the target were separated by one and eight electrodes for each sequence duration.

2.2.4 Discussion

Experiment 1 investigated if electrode separation promotes voluntary stream segregation and whether a segregated percept needs time to build up in a segregation-promoting paradigm. The detection performance was assumed to improve if the listeners would perceptually segregate the A and B sequences. Thus, greater d' scores represent higher likelihood for a segregated percept.

Earlier studies that considered temporal tasks to assess streaming abilities of CI listeners reported a large variability in their results (Cooper and Roberts, 2009; Hong and Turner, 2006; Tejani et al., 2017). Such variability is likely to represent differences in both streaming abilities as well as temporal discrimination abilities across subjects. In an attempt to minimize the variability due to individual differences in temporal discrimination abilities, Δt was adjusted for each listener. Despite this individual adjustment of the task difficulty, the results still varied considerably across listeners (figure 2.2 and 2.3).

Greater d' scores were observed, overall, for the large than for the small electrode separation between the target and the distractor stream. Thus, a large electrode separation facilitated the detection task, suggesting that CI listeners were able to make use of place cues to segregate the A and B sequences. This finding is consistent with reports from previous studies (e.g. Chatterjee et al., 2006; Hong and Turner, 2006; Tejani et al., 2017). However, this was only observed for the long sequence and not for the short one, in which d' scores did not depend on the electrode separation (figure 2.3, panels A and C). The build-up process of a two-stream percept has been widely reported for NH listeners, both in obligatory (Roberts et al., 2008; Thompson et al., 2011) and voluntary stream segregation (Micheyl et al., 2005; Nie and Nelson, 2015). Presumably, the short sequence in the present study was not long enough to allow such build-up process to occur in the CI listeners. The results from the no-distractor condition demonstrated that detecting the delay on the B sequence per se was not affected by the sequence duration. Thus, the greater d' scores achieved for the large rather than for the short electrode separation in the long sequence, are likely to represent the build-up of a two stream percept. The results from experiment 1 suggest that a similar build-up process is experienced by both NH and CI listeners during voluntary stream segregation. This is consistent with the findings from Böckmann-Barthel et al. (2014), who investigated the time course of stream segregation in CI listeners as a function of frequency separation and found similar trends in CI and NH listeners. In that study, the listeners directly reported their percept without any specific instructions encouraging integration or segregation of the sounds. Thus, it is possible that the reports from the CI listeners reflected pitch or electrode discrimination instead of stream segregation (Chatterjee et al., 2006; Cooper and Roberts, 2007). Such uncertainty was avoided in the present study by using a detection task that specifically promotes segregation.

Cooper and Roberts (2009) did not find an effect of electrode separation on voluntary stream segregation performance in CI listeners. However, their sequences had a fixed duration of 2.2 seconds. This is longer than the short sequence (1.24 s) and shorter than the long sequence (3.96 s) used in the present study. Thus, the results from the present study suggest that CI listeners need between about 1.2 and 4 s to build up a two-stream percept on a segregation-promoting paradigm. In the study of Cooper and Roberts (2009), such build-up effect could have been significantly reduced by introducing large loudness differences between the target and the distractor sounds, which has been shown to be a strong cue for stream segregation in CI listeners (Marozeau et al., 2013). In their study, CI listeners performed near-chance level in the absence of loudness cues, but could segregate the target sounds when the distractor sounds were attenuated by at least 50% of the listener's dynamic range. In the present study, CI listeners required shorter Δt values in order to avoid ceiling effects in the absence of the distractor stream. This implies that performance in the rhythm detection task was substantially affected by the presence of a distractor stream even when the electrode separation between the target and the distractor was as large as eight electrodes. Thus, even though CI listeners seem to be able to achieve a segregated percept and exhibit a similar build-up process as the one reported for NH listeners, it is likely that they need longer time to achieve a fully segregated percept when only place cues are provided.

The results reported in the present study are similar to the ones obtained by Nie and Nelson (2015) who used a similar segregation-promoting paradigm to investigate the effect of spectral separation and sequence duration on stream segregation in NH listeners. They found a significant interaction between the sequence duration and the spectral separation between the A and B sounds. A corresponding interaction between electrode separation and sequence duration was found here for CI listeners. Tejani et al. (2017) made use of the irregular rhythm detection task (Cusack and Roberts, 2000) to assess obligatory stream segregation abilities of both NH and CI listeners. Despite the variability of the CI group, the results showed similar trends for both NH and CI listeners, with no significant differences between the groups. The similarity in the trends observed in both groups supports the idea that CI listeners and NH listeners might experience both voluntary and obligatory stream segregation in a similar way.

2.3 Experiment 2: Estimating the fission boundary

2.3.1 Rationale

Experiment 2 investigated how large the electrode separation needs to be for segregation to occur with a subset of the listeners from experiment 1. The same rhythm detection task as in experiment 1 was used. However, the sequence duration was fixed at 3.96 s (long sequence in experiment 1) and only the electrode separation was varied.

The rhythm detection task used in experiments 1 and 2 encouraged the listeners to focus on the temporally regular B sounds and ignore the jittered A sounds. However, the distribution of possible onset-to-onset gaps between the last A and B sounds was shifted by $+\Delta t$ in the delayed with respect to the non-delayed sequences, providing the listeners with an extra cue to perform the task. As a result, listeners could have achieved above chance performance even if they would have been unable to segregate the sounds. Due to the individual adjustment of Δt, the d' reflecting chance level performance varied across listeners. In experiment 2, an ideal observer model was used to establish the upper limit of performance for each listener when the streams were assumed to be perceived as fused.

2.3.2 Methods

The methods used in experiment 2 are identical to those used in experiment 1, unless otherwise stated.

Listeners

Six CI listeners (four female and two male) participated in this experiment. The listeners were 21 to 74 years old (mean: 52 years, SD: 23 years; see table 2.1) and had no residual hearing in their implanted ear.

Stimuli and conditions

Five electrode separations were tested with a sequence duration of 12 AB pairs, leading to a nominal duration of 3.96 s. The target stream (B) was always played through electrode 11 and the distractor stream (A) was played through either electrode 11 (*no-difference* condition), 12, 14, 16 or 19 depending on the condition. Each pulse burst was presented at a rate of 900 pps. As in experiment 1, an additional delay (Δt) was sometimes added before the last burst of the target stream (see table 2.2).

Procedure

The first session comprised a brief explanation of the task, the loudness balancing and some of the blocks from the rhythm detection task. The remaining blocks as well as the *no-difference* condition were tested in the second session. A total of ten different sequences were presented to the listeners, resulting from the combination of five possible distractor electrodes (11, 12, 14, 16 or 19) and two different Δt values (delayed or non-delayed). The distractor electrodes 12, 14, 16 and 19 were presented in pseudorandom order, ensuring that different distractor electrodes were presented in consecutive sequences. The *no-difference* condition was tested on a separate block.

Ideal observer model

An ideal observer model was used to simulate the best possible performance that listeners could achieve if the delay between the last A and B sounds would be the only available cue. The model categorized individual trials as delayed or non-delayed by evaluating the gap between the last A and B sounds of a given sequence and comparing it to the nominal gap between consecutive A and B sounds (i.e. 170 ms). A given trial was categorized as delayed if the gap between the last A and B sounds was larger than the nominal gap. Otherwise, the trial was categorized as non-delayed. Because Δt values were adjusted individually, the probability of giving a correct answer when fusing the A and B streams (chance level) was different for each listener. Thus, the gap between the last A and B sounds of each presentation was stored for each listener and condition and used as input to the IO model. The IO model generated a d' estimate for each listener and condition. Segregation was considered to occur when CI listeners' performance was significantly better than the one achieved by the IO model.

Statistical alanlysis

A mixed-effects linear model was fitted to the d' scores. The electrode separation and data type (listener's data or IO model prediction) were treated as fixed effects terms whereas listener-related effects were treated as random effects (random intercept and slope). Statistical contrasts between the individual listener's data and their respective IO model predictions were performed using t-tests with the mean and standard error from each d' estimate. The resulting p-values were adjusted for multiple comparisons for controlling the false discovery rate (Benjamini and Hochberg, 1995).

2.3.3 Results

Figure 2.4 shows the individual results from experiment 2. For each listener, the d' scores are shown for each target-distractor electrode separation. The experimental data are indicated by the black filled circles. Estimates from the IO model are indicated by the gray triangles. Adjusted p-values resulting from the statistical contrast between the achieved d' scores and the IO model predictions are indicated with asterisks.

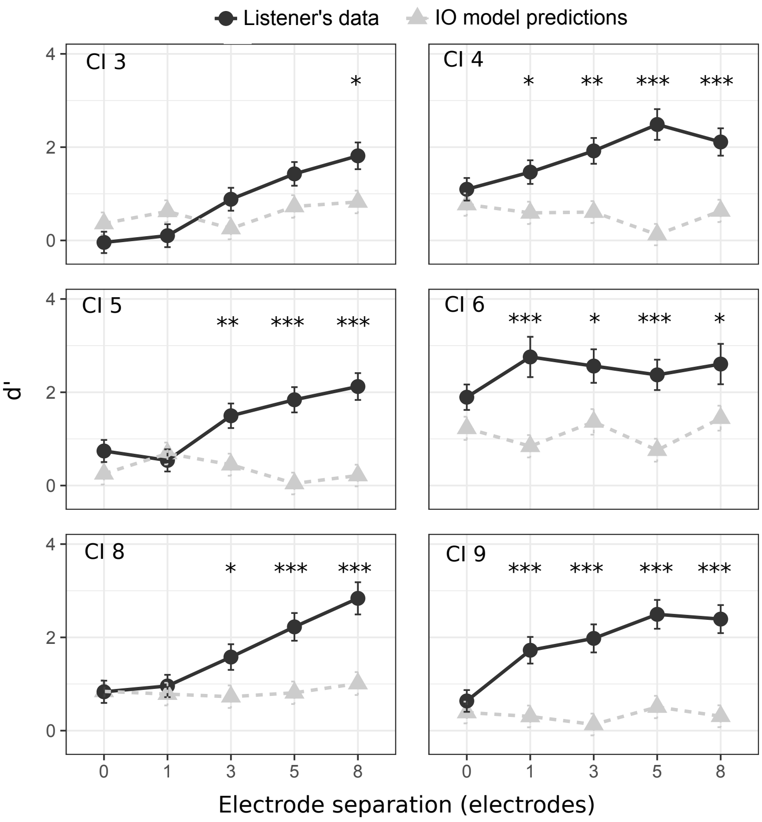


Figure 2.4: Individual sensitivity (d') scores to the delayed tone for each electrode separation (black circles, solid line) as well as the corresponding ideal observer model prediction (gray triangles, dashed line). Error bars represent the standard errors of the d' estimates. A statistically significant difference between the IO model predictions and the listener's data is indicated by one asterisk if 0.05>p>0.01, two asterisks if 0.01>p>0.001 and three asterisks if p<0.001.

Sensitivity scores generally increased for larger electrode separations between the A and B sequences. On average, listeners required a minimum separation of 2.8 electrodes to obtain significantly larger d' scores than those from the IO model. However, a large variability was observed across listeners: while listeners 4, 6 and 9 obtained significantly larger d' scores than the IO model when the A and B sequences were separated by one or more electrodes, listeners 5 and 8 required a minimum separation of 3 electrodes and listener 3 could only obtain larger d' scores than the IO model for a separation of 8 electrodes. None of the listeners achieved significantly larger d' scores than those predicted by the IO model (figure 2.4) for the *no-difference* condition.

Figure 2.5 contains d' scores from all listeners (dark gray boxes) and the corresponding IO model estimates (light gray boxes). The d' scores from the listeners increased monotonically with the electrode separation between the A and B sounds, possibly reaching a plateau at a separation of five electrodes. The IO model predictions were rather constant across the different electrode separations, although they showed some variability both across the different electrode separations and across listeners. The variability across electrode separations reflects the limited number of observations used for calculating the d' (60 observations of the delayed and 60 observations of the non-delayed sequences) since the IO model predictions were solely based on the gap between the last A and B sounds and did not depend on the electrode separation between the A and B sequences. The variability across listeners reflects the use of individual Δt values in this experiment. The IO model predictions were related to Δt, since larger Δt values increased the difference between the distributions of possible gaps between the last A and B sounds of the delayed and non-delayed sequences. Thus, the larger was Δt, the larger was the predicted d'.

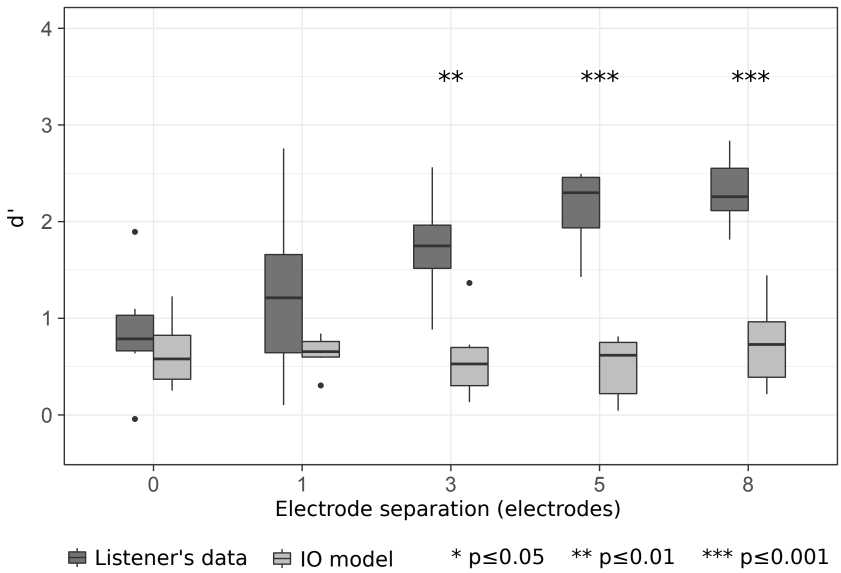


Figure 2.5: Sensitivity (d') scores to the delayed tone for each electrode separation. Data from the CI listeners is plotted in dark gray and predictions from the ideal observer model in light gray.

Table 2.3 summarizes the results from the statistical contrast between d' scores obtained by CI listeners and those predicted by the IO model, based on the lsmeans estimates obtained from the mixed-effects linear model. Both electrode separation [$F(4, 40) = 12.810$, $p < 0.001$], data type [$F(1, 5) = 33.496$, p = 0.002] and their interaction [$F(4, 40) = 13.083$, $p < 0.001$] were found to be significant factors in the statistical model. Listeners' d' scores were significantly larger than those obtained with the IO model for a separation of 3 or more electrodes between the A and B streams.

Table 2.3: Summary of the statistical contrast between d' scores obtained by CI listeners and IO model for each electrode separation. Statistical contrasts were performed on the lsmeans estimates from the mixed-effects linear model for each data type and electrode separation (df = 13.83).

Electrode separation	Difference estimate	t ratio	p-value
0	0.220	0.932	0.992
1	0.619	2.627	0.290
3	1.147	4.870	0.007
5	1.645	6.984	<0.001
8	1.574	6.683	<0.001

2.3.4 Discussion

Experiment 2 combined measurements of performance using the rhythm detection task from experiment 1 and predictions from an IO model to estimate the minimum electrode separation needed to segregate the streams. Consistent with the results from experiment 1, greater d' scores were achieved for larger electrode separations, demonstrating that a larger electrode separation facilitated the segregation of the sounds. Moreover, all CI listeners achieved significantly greater d' scores than those predicted by the IO model, indicating that all listeners were able to achieve a segregated percept.

Böckmann-Barthel et al. (2014) made use of direct reports of perception from CI listeners to assess the role of place cues on stream segregation. Even though their study did not aim to estimate the fission boundary, they reported an ambiguous percept for a frequency separation of six semitones between the A and the B sounds and suggested that a separation of two to three electrodes might be needed by CI listeners to segregate the sounds. With a similar paradigm as Böckmann-Barthel et al. (2014), Chatterjee et al. (2006) and Cooper and Roberts (2007) found little or no evidence for ambiguous percepts. The results from these studies indicated the proportion of time where listeners reported a two-stream percept and thus, they might be influenced by how listeners were instructed to perform the task. Ultimately, if listeners are uncertain about what to listen for, they might report pitch or electrode discrimination instead of segregation (Chatterjee et al., 2006; Cooper and Roberts, 2007). The results from experiment 2, where the fission boundary was assessed through a rhythm detection task, support the hypothesis of Böckmann-Barthel et al. (2014). Three out of six listeners were able to segregate the sounds when they were presented from adjacent electrodes and all (but one) listeners were able to experience a segregated percept with a separation of three electrodes.

Temporal perception has been found to be similar in CI and NH listeners (e.g. Moore and Glasberg, 1988; Shannon, 1989, 1992) and previous studies have demonstrated that CI listeners are able to make use of temporal cues to segregate sounds (e.g. Duran et al., 2012; Hong and Turner, 2009). Temporal regularity and predictive processing are known to influence the representation of auditory objects, with irregular sounds being more likely to be segregated (for a review, see Bendixen, 2014). In the present study, the distractor stream was always temporally irregular, possibly contributing to the segregation process. The temporal regularity properties of the streams were kept constant across conditions; therefore, it cannot account for the improvement in performance observed for the larger electrode separation. Nie et al. (2014) observed that listeners were able to segregate sequential sounds under attentive listening even when only temporal regularity cues were present. In the present study, none of the CI listeners achieved significantly larger d' scores than those predicted by the IO model for the *no-difference* condition, where both A and B streams were presented through the same electrode and with identical pulse rate. Thus, even though the temporal irregularity of the distractor stream may contribute to the segregation process both in NH and in CI listeners (e.g. Nie et al., 2014; Rajendran et al., 2013), in the present study, this cue was not found to be strong enough to elicit a segregated percept. Instead, place cues were the dominant cue used by CI listeners to segregate the streams.

2.4 Summary and conclusion

The present study assessed the effect of place cues on voluntary stream segregation in CI listeners. The results from experiment 1 suggest that CI listeners can make use of place cues to voluntarily segregate sounds. Moreover, a build-up process similar to that reported in NH listeners was observed. In experiment 2, all (but one) listeners were able to segregate the sounds for electrode separations of three electrodes, with some listeners being able to segregate sounds coming from adjacent electrodes. Experiment 2 also validated the use of the rhythm detection task to assess the effect of electrode separation on stream segregation in the presence of temporal regularity cues, since temporal regularity was not salient enough to elicit a segregated percept in the absence of place cues. Altogether, place cues seem to play an important role for the segregation of sounds, allowing CI listeners to segregate sequentially presented sounds. However, these findings are based on a relatively simple paradigm and should not be extrapolated to more complex and realistic scenarios without further investigation. It is possible that the limitations experienced by CI listeners in complex listening scenarios, such as speech intelligibility in a noisy environment, arise from the degraded frequency resolution. Current sound coding strategies result in a wide range of electrodes being active most of the time which might limit the place information available to the listener (e.g. Tejani et al., 2017).

Acknowledgments

We would like to thank all volunteers who participated in this study. We thank the two anonymous reviewers for the helpful and constructive comments on an earlier version of the manuscript, Per B. Brockhoff and Alexandra Kuznetsova for their help and guidance with the statistical analysis of the data and our colleagues from the Hearing Systems Group for valuable comments and stimulating discussions. This work was supported by the Oticon Centre of Excellence for Hearing and Speech Sciences (CHeSS) and the Carlsberg Foundation. The research equipment was provided by Cochlear Ltd.

3

The role of temporal cues in voluntary stream segregation [c]

Abstract

The role of temporal cues in sequential stream segregation was investigated in cochlear implant (CI) listeners using a delay detection task composed of a sequence of bursts of pulses (B) on a single electrode interleaved with a second sequence (A) presented on the same electrode with a different pulse rate. In half of the trials, a delay was added to the last burst of the otherwise regular B sequence and the listeners were asked to detect this delay. As a jitter was added to the period between consecutive A bursts, time judgments between the A and B sequences provided an unreliable cue to perform the task. Thus, the segregation of the A and B sequences should improve performance. The pulse rate difference and the duration of the sequences were varied between trials. The performance in the detection task improved by increasing both pulse rate differences and sequence duration. This suggests that CI listeners can use pulse rate differences to segregate sequential sounds and that a segregated percept builds up over time. In addition, the contribution of place vs temporal cues for voluntary stream segregation was assessed by combining the results from the present study with those from our previous study, where the same paradigm was used to determine the role of place cues on stream segregation. Pitch height differences between the streams accounted for the results from both studies, suggesting that stream segregation is related to the salience of the perceptual difference between the sounds.

[c] This chapter is based on Paredes-Gallardo et al. (2018b,c).

3.1 Introduction

Auditory stream segregation has often been investigated using an *auditory streaming* paradigm (e.g. Bregman, 1990; Carlyon, 2004; Moore and Gockel, 2002, 2012). In this paradigm, two repeating sounds (A and B), typically two pure tones with different frequencies, are presented sequentially to the listener who might integrate them into a single stream or segregate them into two separate streams. Whereas large frequency differences between the sounds facilitate segregation, small frequency differences promote integration (e.g. Bregman and Campbell, 1971; Van Noorden, 1975). In normal-hearing (NH) listeners, stream segregation is also influenced by other stimulus properties than frequency differences, such as differences in the temporal envelope (e.g. Cusack and Roberts, 2000; Grimault et al., 2002; Iverson, 1995; Singh and Bregman, 1997; Vliegen and Oxenham, 1999), the phase spectrum (e.g. Roberts et al., 2002) or the spatial characteristics (e.g. David et al., 2015; Sach and Bailey, 2004; Stainsby et al., 2011). Therefore, it has been hypothesized that sequential stream segregation may be directly related to the degree of the perceptual difference between the sounds (Moore and Gockel, 2002, 2012).

In electric hearing, perceptual differences can be elicited by varying the electrode (place cues) or the pulse rate (temporal cues) of the stimulation (e.g. Eddington et al., 1978; Lamping et al., 2018; Landsberger et al., 2016; Shannon, 1983). Both electrode and pulse rate of stimulation can contribute to the perception of pitch height. The stimulation of apical electrodes and the use of low pulse rates are generally associated with a lower pitch percept than the stimulation of basal electrodes and the use of a high pulse rate (e.g Lamping et al., 2018; Landsberger et al., 2016). It has been suggested that cochlear implant (CI) listeners might be able to combine place and rate information (e.g. Luo et al., 2012; McKay et al., 2000; Rader et al., 2016). However, McKay et al. (2000) reported no advantage of consistent combinations of place and rate information (e.g. a slow pulse rate stimulating an apical electrode) over inconsistent combinations (e.g. a slow pulse rate stimulating a basal electrode) for the discrimination of sounds. Thus, place and temporal cues are considered to be perceptually orthogonal and independent cues in electric hearing (e.g. Marimuthu et al., 2016; McKay et al., 2000; Tong et al., 1983).

Previous studies investigating auditory stream segregation in CI listeners have focused on the role of place cues. In contrast to the results from studies in

NH listeners, some of these studies did not observe any effect of the sequence duration (i.e. build-up) or the tone presentation rate on the ability to segregate the sounds (Chatterjee et al., 2006; Cooper and Roberts, 2007, 2009). Nevertheless, other studies found results consistent with those from studies in NH listeners, suggesting that there are circumstances in which CI listeners can use place cues to segregate sounds (Böckmann-Barthel et al., 2014; Chatterjee et al., 2006; Hong and Turner, 2006; Paredes-Gallardo et al., 2018a; Tejani et al., 2017). Moreover, the results from several studies suggest that CI listeners need time to build up a segregated percept (Böckmann-Barthel et al., 2014; Hong and Turner, 2006; Paredes-Gallardo et al., 2018a), even though the build-up might be slower for CI listeners than for NH listeners (Paredes-Gallardo et al., 2018a).

The effect of temporal cues on stream segregation in CI listeners has been investigated by Chatterjee et al. (2006), Duran et al. (2012), and Hong and Turner (2009). The results from these studies suggest that large differences in the amplitude modulation or the pulse rate between the A and the B sounds facilitate both voluntary stream segregation (Chatterjee et al., 2006; Hong and Turner, 2009) and obligatory stream segregation (Duran et al., 2012). Chatterjee et al. (2006) observed a larger probability of a two-stream percept with increasing sequence duration (i.e. build-up) in one listener. To our knowledge, no other study has investigated whether CI listeners experience a build-up as a function of pulse rate or amplitude modulation differences.

The present study examined the role of temporal cues on voluntary stream segregation in CI listeners. Delay detection performance was measured in a paradigm where the listeners were required to make time judgements between consecutive sounds of a target stream while ignoring a temporally irregular distractor stream. The task became easier if the listeners could segregate the target from the distractor and, thus, the performance in the detection task was affected by the stream segregation ability of the listeners. This paradigm has previously been used to investigate the role of spectral and temporal cues on stream segregation in NH listeners (e.g. Nie and Nelson, 2015; Nie et al., 2014) and the role of place cues in CI listeners (Paredes-Gallardo et al., 2018a). Here, temporal cues were induced by varying the pulse rate at a fixed cochlear location. The aim of the present study was to clarify whether CI listeners can use pulse rate differences (Δrate) to segregate the streams and whether a two-stream percept builds up over time. The fission boundary was estimated as a function of Δrate. Furthermore, the contribution of place vs temporal cues for voluntary stream segregation was assessed by combining the results from the present study with those from Paredes-Gallardo et al. (2018a), presented in chapter 2 from this thesis. Electrode and pulse rate differences were converted to pitch height differences (Δpitch) using data from a verbal attribute magnitude estimate experiment Lamping et al., 2018. If stream segregation is related to the salience of the perceptual difference between the sounds, the Δpitch between the target and the distractor stream should account for the results from both studies.

3.2 Methods

3.2.1 Listeners

Seven CI listeners (six female and one male) participated in this experiment. The listeners were aged between 19 and 74 years (mean: 50.8 years, SD: 21.5 years; see table 3.1), had no residual hearing and were bilateral CI users. All listeners were users of the Cochlear Ltd. (Sydney, Australia) implant. Six of the listeners had previously participated in the study presented in chapter 2, where the same paradigm was used to assess the effect of place cues on stream segregation. The same listener IDs are used in both chapters. All listeners provided informed consent prior to the study and all experiments were approved by the Science-Ethics Committee for the Capital Region of Denmark (reference H-16036391).

Table 3.1: Relevant information about CI listeners. CI24RE and CI24R are two implant models, also known as Nucleus-24 and Freedom, respectively.

ID	Age	Gender	Onset of deafness	Implant model (ear)	Years of experience
CI 1	19	F	Prelingual	CI24RE (right)	16
CI 4	74	F	Postlingual	CI24R (left)	13
CI 5	73	M	Postlingual	CI24RE (right)	3
CI 6	64	F	Perilingual	CI24R (right)	15
CI 8	61	F	Perilingual	CI24RE (right)	3
CI 9	21	F	Prelingual	CI24RE (left)	16
CI 10	44	F	Prelingual	CI24RE (left)	5

3.2.2 Stimuli and conditions

The stimulation paradigm is illustrated in figure 3.1, where the different panels represent the different conditions. A sequence of 50 ms bursts of pulses (B) presented on a single electrode was interleaved with a sequence (A) presented on the same electrode with a different pulse rate. In half of the trials, a small delay (Δt) was added to the last burst of the otherwise regular B sequence (the target stream). The listeners were asked to indicate after each trial whether or not the last sound of the sequence was delayed. The nominal onset-to-onset interval between consecutive B sounds was 340 ms, and a random jitter was added to the onset-to-onset interval between consecutive A sounds. The duration of the jitter applied to each A sound was drawn from a rectangular distribution with a range of ±110 ms. Thus, the onset-to-onset interval between the A and B sounds was 170 ms ± jitter, as illustrated in figure 3.2. Consecutive A and B sounds were always separated by a minimum interval of 10 ms. The temporal irregularity of the distractor stream made across-streams time judgements an unreliable cue to perform the task. Therefore, to optimize performance, the listeners needed to compare the time interval between the last two B-sounds with those between previous B-sounds. Thus, the task became easier when the A and B sequences fell into different streams (Micheyl and Oxenham, 2010a; Nie and Nelson, 2015; Nie et al., 2014), encouraging the listener to segregate the streams.

Each A and B sound consisted of a 50-ms burst of biphasic pulses presented at electrode 11[d], located at the mid-point of the array, in monopolar mode. Each biphasic pulse had a phase width of 25 µs and a phase gap of 8 µs. The stimuli were presented through the Nucleus Implant Communicator research interface (NIC v2, Cochlear Limited, Sydney).

The ability of CI listeners to perceive pitch changes as a function of pulse rate (temporal pitch) has been reported to be limited to rates below 300/400 pps (e.g. Shannon, 1983; Tong and Clark, 1985; Townshend et al., 1987). In the present study, the target stream was played with a constant rate of 300 pps, while the A sequence was played with a lower pulse rate of either 80, 140, 200 or 260 pps, leading to a Δrate between the streams of 220, 160, 100 or 40 pps depending on the condition.

[d] In the Cochlear electrode array, electrode 1 is the most basal electrode and electrode 22 the most apical one.

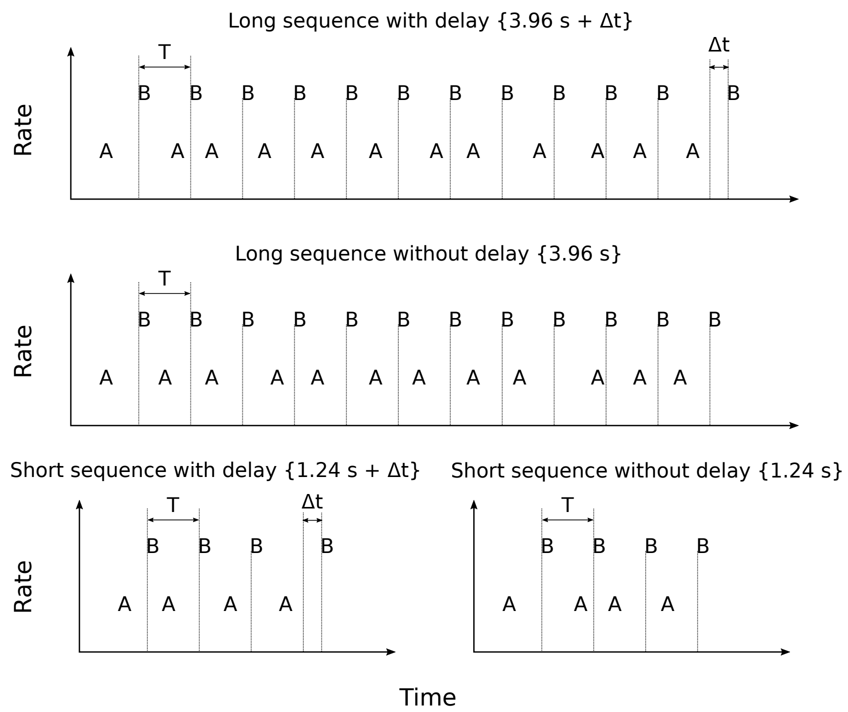


Figure 3.1: Graphical representation of the experimental paradigm. T represents the onset-to-onset interval and Δt is the delay of the last B sound. The long sequence, with and without Δt, is shown in the upper and middle panels. The short sequence, with and without Δt, is illustrated in the lower left and lower right panels, respectively. The rate difference between A and B sounds varied across conditions.

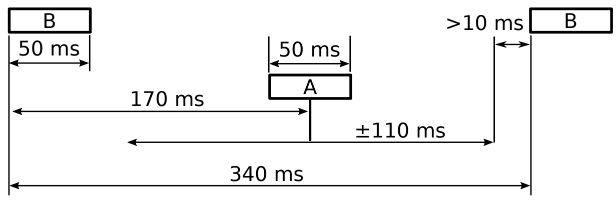


Figure 3.2: Graphical representation of the timing between the B and the A sounds.

All sequences started with the distractor stream (A) and ended with the target stream (B), as illustrated in Figure 1. Two sequence durations were tested. The long sequence consisted of 12 AB pairs (3.96 s without Δt) and the short sequence consisted of 4 AB pairs (1.24 s without Δt). Performance in the detection task for the long and the short sequences was also measured without the distractor stream (control conditions). These conditions were easier than the test conditions and, thus, a shorter Δt was used to avoid ceiling effects. A *no-difference* condition (Δrate= 0) was also tested for the long sequence. In this condition, both target and distractor were presented from the same electrode and with the same pulse rate. Both the control and the *no-difference* condition were identical to those described in chapter 2. Thus, only listener 10, who did not participate in the study presented in chapter 2, performed those conditions in the present study. For the remaining listeners, the results from the control and the *no-difference* condition were obtained from chapter 2.

For each combination of Δrate and sequence duration, 60 presentations of the delayed sequence and 60 presentations of the non-delayed sequence were used to calculate the listener's sensitivity (d') to the delayed target.

3.2.3 Loudness balancing

Loudness has been found to be an effective cue for sound segregation in CI listeners (e.g. Cooper and Roberts, 2009; Marozeau et al., 2013). The stimuli were therefore loudness-balanced in the present study. Categorical loudness scaling was used to find the most comfortable levels (MCL) for each listener and stimulus. Each pair of target and distractor sounds was then loudness matched by the listeners using the procedure described in chapter 2. The loudness matching was performed in the beginning of each session. The level of the loudness balanced stimuli did not markedly change between sessions.

3.2.4 Delay adjustment procedure

Individual Δt values were chosen such that listeners would be equally sensitive to the delayed target in a given condition, minimizing the effect of individual differences on the detection performance in the auditory streaming task. To facilitate the comparison of the results from the present study and those from the study presented in chapter 2, the same individual Δt values were used in the two studies. For Listener 10, who did not participate in the study presented in chapter 2, Δt was derived using the criterion described in chapter 2: Δt was defined as the delay leading to d' = 2 for the long sequence whereby the 50 ms bursts of pulses were presented at 900 pps to electrodes 11 (A) and 19 (B) (table 3.2).

The individual Δt to be used in the control condition was also derived as in chapter 2, i.e. the Δt corresponding to d' = 3 for the long sequence without the distractor stream. This d' value was chosen to keep the control conditions relatively easy while avoiding ceiling effects.

Table 3.2: Individual Δt values as obtained from the delay adjustment procedure.

ID	Δt [ms] for $d' = 2$	Δt [ms] for control condition, $d' = 3$
CI 1	40	30
CI 4	45	35
CI 5	35	32
CI 6	80	55
CI 8	60	28
CI 9	35	30
CI 9	60	40

3.2.5 Procedure

The experiments took place in a double-walled sound-attenuating booth and were organized into two sessions, each lasting 2h including short breaks. For listener 10, the first session included a brief description of the task and the delay adjustment procedure, as well as a 10 – 15 min training on the detection task. The other listeners had participated in the study presented in chapter 2, and were therefore familiar with the paradigm.

A one-interval two-alternative forced-choice procedure was used, where the listeners were asked to report if the last target sound of the sequence was delayed or not. A one interval task was chosen to minimize the attentional effort required to perform the task (Nie and Nelson, 2015). The sequences were organized in twelve blocks, six with long sequences and six with short sequences, presented in random order. On a given block, the four Δrate conditions were presented in pseudorandom order, ensuring that the same Δrate condition would not be presented in consecutive sequences. Thus, the first sound of each sequence always had a different rate, contributing to resetting the build-up of a two-stream percept after each presentation (Roberts et al., 2008). Each Δrate condition was presented 20 times in each block (10 delayed and 10 non-delayed presentations). The *no-difference* condition was tested in a separate block.

The control conditions were tested in four blocks (two with long sequences and two with short sequences), with each block containing 30 repetitions of the delayed and 30 repetitions of the non-delayed sequences. The control blocks were randomly presented at the beginning or at the end of either session.

3.2.6 Ideal observer model

The distribution of possible onset-to-onset gaps between the last A and B sounds was different in the delayed and the non-delayed sequences. The gap between the last A and B sounds in the delayed sequence was, on average, Δt ms longer than the one in the non-delayed sequence. Therefore, the listeners had an extra cue, proportional to Δt, to perform the task. As in chapter 2, an ideal observer (IO) model was used to simulate the best possible performance that each listener could achieve if the gap between the last A and B sounds would be the only available cue. The model categorized individual trials as delayed or non-delayed by evaluating the gap between the last A and B sounds of a given sequence and comparing it to the nominal gap between consecutive A and B sounds (i.e. the gap of 170 ms, when no jitter has been applied). A given trial was categorized as delayed if the gap between the last A and B sounds was larger than the nominal gap. Otherwise, the trial was categorized as non-delayed. Because Δt was adjusted individually, the probability of a correct response when fusing the A and B streams (chance level) was different for each listener. Thus, the gap between the last A and B sounds of each presentation, listener and condition was used as input to the IO model. The IO model generated a d' estimate for each listener and condition. Segregation was considered to occur when the CI listeners' performance was significantly better than the one predicted by the IO model.

3.2.7 Statistical analysis

Unless otherwise specified, statistical inference was performed by fitting a mixed-effects linear model to the computed d' scores. The experimental variables and their interactions were treated as fixed effects whereas listener-related effects were treated as random effects with random intercepts and slopes. The Δrate values were calculated from the log-transformed rate and were back-transformed after the post-hoc analysis for an easier interpretation of the results. The model was implemented in `R` using the `lme4` library (Bates et al., 2014) and the model selection was carried out with the `lmerTest` library (Kuznetsova et al., 2017) following the backwards selection approach based on step-wise deletion of model terms with high p-values (Kuznetsova et al., 2015). The p-values for the fixed effects were calculated from F-tests based on Sattethwaite's approximation of denominator degrees of freedom and the p-values for the random effects were calculated based on likelihood ratio tests (Kuznetsova et al., 2015). The post-hoc analysis was performed through contrasts of least-square means using the `lsmeans` library (Lenth, 2016) and the `lme4` model object. The p-values were corrected for multiple comparisons using the Tukey method.

Statistical contrasts between the individual listeners' data and their respective IO model predictions were performed using t-tests with the mean and standard error from each d' estimate. The resulting p-values were adjusted for multiple comparisons controlling for the false discovery rate (Benjamini and Hochberg, 1995).

3.3 Results

The individual results are shown in figure 3.3, where each row represents the results for an individual listener. Sensitivity scores for the short and long sequences are shown in the left and right columns, respectively. The experimental data are indicated by the blue circles. Estimates from the IO model are indicated by the green triangles. Statistically significant differences between the achieved d' scores and the IO model predictions are indicated by the asterisks.

For the long sequence (right column), d' scores generally increased with increasing Δrate. For the largest Δrate condition, all listeners achieved larger d' scores than the IO model. In contrast, for the *no-difference* condition (Δrate$= 0$), none of the listeners achieved significantly larger d' scores than the IO model. A large across-listener variability was observed, with some listeners exhibiting little or no improvement in detection performance towards larger Δrate values. For the short sequence (left column), no general trend was observed in the d' scores with increasing Δrate. For some listeners, d' scores did not significantly increase with increasing Δrate (i.e. listeners 1, 5, 8 and 9). For three of the seven listeners (i.e. listeners 4, 6 and 10), d' scores increased with increasing Δrate and were larger than the IO model predictions for the largest Δrate condition (i.e. 220 pps).

Figure 3.4 shows the d' scores for all listeners and conditions. The results from the short and long sequences are shown in separate panels. The results for the control (no distractor) condition for the short and the long sequences are shown in the right-most panel. The lsmeans estimates and the 95% confidence interval from the statistical model fitted to the data are represented by solid and dashed lines, respectively. The data from the listeners and the predictions from

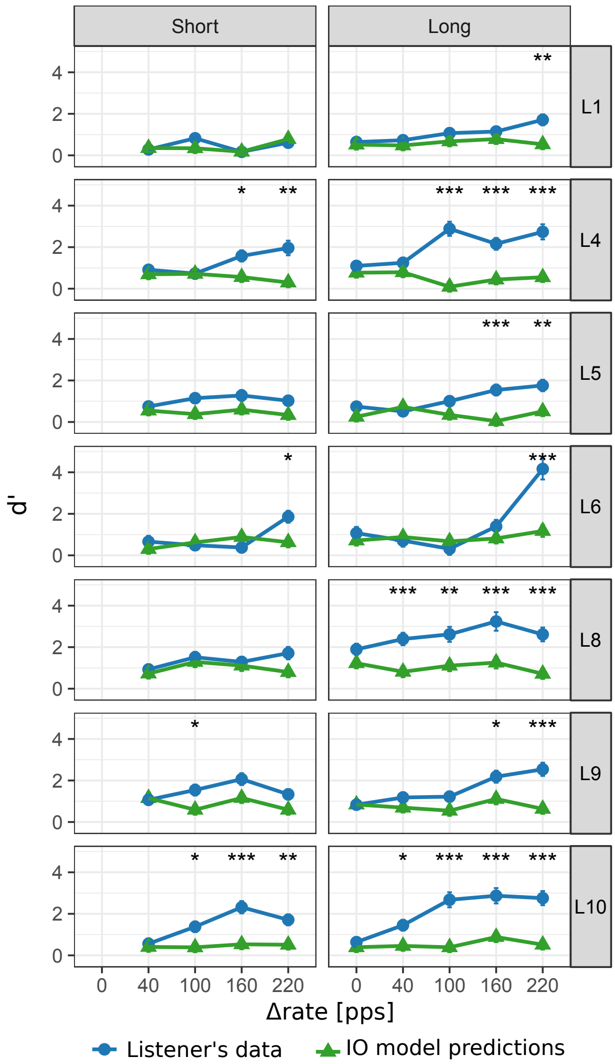


Figure 3.3: Individual sensitivity (d') scores to the delayed B-sound for each Δrate and sequence duration (blue circles) as well as the corresponding ideal observer model prediction (green triangles). Error bars represent the standard error of the d' estimates. The error bars often fall within the symbols and are therefore not always visible in the graph. A statistically significant difference between the IO model predictions and the listener's data is indicated by one asterisk if $0.05>p>0.01$, two asterisks if $0.01>p>0.001$ and three asterisks if $p<0.001$.

the IO model are represented by boxes. Different colors represent the measured data and the IO model predictions.

The sensitivity scores (d') increased with Δrate [$F(1, 29.95) = 23.051$, $p < 0.001$]. The main effect of sequence duration [$F(1, 98.70) = 1.990$, $p = 0.162$] and of the data type (listener's vs IO model) [$F(1, 11.53) = 2.526$, $p = 0.139$] were found to be non-significant. However, a significant interaction was found

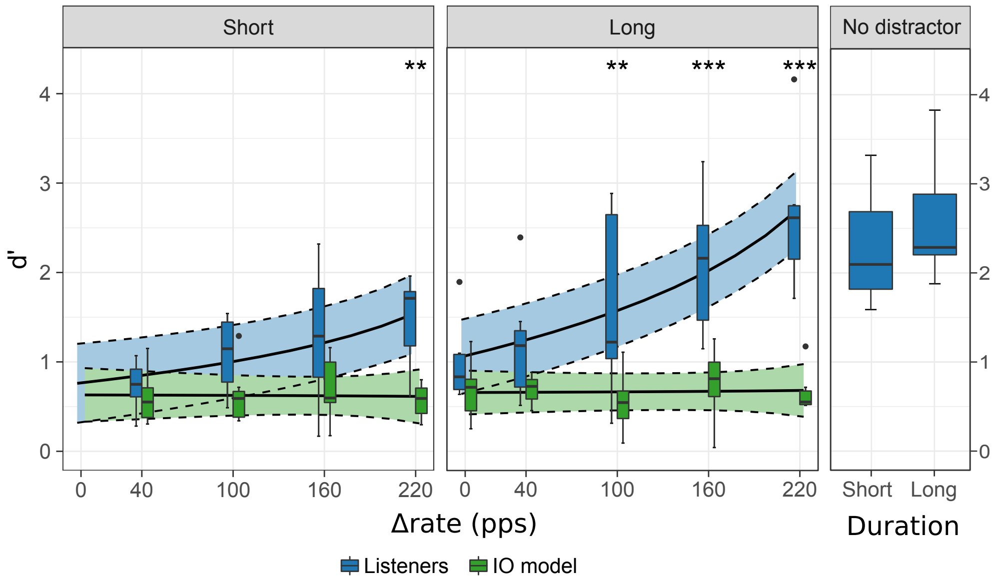


Figure 3.4: Sensitivity (d') scores to the delayed B-sound for each Δrate and sequence duration. The control condition (no distractor) for both long and short sequences is shown in the right-most panel. The boxes illustrate data from the listeners (blue) and the corresponding IO model predictions (green). The solid lines represent the lsmeans estimate from the statistical model. Its 95% confidence interval is indicated with dashed lines and the corresponding shaded area. A statistically significant difference between the IO model predictions and the listener's data is indicated by one asterisk if 0.05>p>0.01, two asterisks if 0.01>p>0.001 and three asterisks if p<0.001.

between Δrate and sequence duration [$F(1, 93.11) = 5.277$, $p = 0.024$], Δrate and data type [$F(1, 80.75) = 38.958$, $p < 0.001$] and Δrate, sequence duration and data type [$F(1, 80.75) = 4.609$, $p = 0.035$].

The increase in the d' scores obtained by the listeners with increasing Δrate was significantly steeper for the long sequence than for the short sequence [$t(86.47) = 3.134$, $p = 0.012$], indicating a greater effect of Δrate for the long than for the short sequences. For the long sequences, the increase of the d' scores with increasing Δrate was significantly steeper for the listeners than for the IO model predictions [$t(79.65) = 6.557$, $p < 0.001$]. The listeners performed significantly better than the IO model for the Δrate values of 100 pps [$t(8.60) = 5.784$, $p = 0.006$], 160 pps [$t(9.63) = 8.354$, $p < 0.001$] and 220 pps [$t(27.59) = 9.454$, $p < 0.001$]. Thus, for the long sequence, the smallest Δrate at which the listeners could segregate the streams (i.e. the fission boundary) was 100 pps (50% relative to the distractor pulse rate). For the short sequences, the increase of the d' scores with increasing Δrate was only marginally steeper for the listeners than for the IO model predictions [$t(79.65) = 2.664$, $p = 0.045$]. The listeners achieved significantly larger d' scores than those from the IO model only for the largest Δrate condition (Δrate= 220 pps) [$t(29.62) = 4.214$, $p = 0.007$, difference estimate = 0.915]. Thus, for the short sequences, the fission boundary was 220 pps (275% relative to the distractor pulse rate).

A paired t-test revealed no significant difference between the d' scores achieved for the long and short sequences in the control condition (no distractor) [$t(6) = 1.515$, $p = 0.180$].

In summary, the performance in the delay detection task improved with increasing Δrate. A larger effect of Δrate was observed for the long than for the short sequence, indicating the build-up of stream segregation.

3.4 Discussion

In the present study, a delay detection task was used to assess the stream segregation abilities of CI listeners. The task became easier when the listeners could segregate the sounds - hence, larger d' scores were achieved in conditions facilitating a segregated percept. Segregation was considered to occur when the d' scores achieved by the CI listeners were significantly larger than those predicted by the IO model.

3.4.1 The role of pulse rate differences in stream segregation

The d' scores obtained by the listeners increased with increasing Δrate, both for the long and for the short sequences, suggesting that the listeners were able to use pulse rate differences to segregate the streams. These findings are consistent with earlier work suggesting that larger differences between the temporal envelopes of the A and the B sounds facilitate a segregated percept both in NH listeners (e.g. Grimault et al., 2002; Roberts et al., 2002; Vliegen and Oxenham, 1999; Vliegen et al., 1999) and in CI listeners (Chatterjee et al., 2006; Duran et al., 2012; Hong and Turner, 2009). Hong and Turner (2009) investigated the role of amplitude modulation differences in stream segregation both in NH and in CI listeners. They used a rhythm detection task that became easier when the A and B sounds were perceptually segregated. In their study, both groups of listeners were found to be able to use differences in the temporal envelope of sequential sounds to voluntarily segregate them. Hong and Turner (2009) presented the stimuli through a loudspeaker and the CI listeners used their own speech processor. Therefore, they had only limited control over the exact stimuli delivered to the listeners (as noted by Cooper and Roberts, 2009). The results from the present study support the findings from Hong and Turner (2009). However, in the present study, temporal cues were elicited by directly manipulating the stimulation rate at a fixed cochlear location, bypassing the listener's speech processor such that there was a better control of the signal delivered to the listeners.

It has been suggested that NH listeners might need larger differences to perceptually segregate two stimuli than to discriminate them (Rose and Moore, 2005). Hong and Turner (2009) measured amplitude modulation frequency discrimination thresholds in NH and CI listeners and compared them to the fission boundary obtained as a function of the amplitude modulation frequency difference between two noise bursts. In both groups, larger amplitude modulation frequency differences were needed to segregate the two sounds than to discriminate them. Previous studies assessing the pulse rate difference limen in CI listeners reported a large variability across listeners and a strong dependency of the difference limen on the pulse rate of the reference sound, i.e. the base rate (e.g. Baumann and Nobbe, 2004; Hoesel and Clark, 1997; Townshend et al., 1987; Zeng, 2002). The difference limen was found to increase with increasing base rate, with values of about 10% at a base rate of 100 pps and about 20% for a base rate of 200 pps. Consistent with the findings from Hong and Turner (2009), the results from the present study suggest that CI listeners need larger differences to segregate the sounds than to discriminate them. This was particularly evident for the short sequence, where a pulse rate difference of 275% of the base rate (80 pps) was needed to segregate the sounds.

Duran et al. (2012) also assessed the role of temporal cues in stream segregation by changing the pulse rate at a fixed cochlear location. Their results suggested that CI listeners can use pulse rate differences to segregate sounds. While in the present study the task became easier when the sounds were perceptually segregated (voluntary stream segregation paradigm), in the study by Duran et al. (2012) the task became easier if the sounds were integrated into a single stream (obligatory stream segregation paradigm). Together, these results suggest that CI listeners can use pulse rate differences for both voluntary and obligatory stream segregation of sequential sounds.

3.4.2 The build-up of a two-stream percept

The d' scores obtained by the listeners increased with increasing Δrate. The effect of Δrate was found to be dependent on the sequence duration, with a steeper growth of d' with increasing Δrate for the long than for the short sequence. Given that in the absence of the distractor stream the performance was not affected by the duration of the sequence, as demonstrated by the results from the *no distractor* condition, these findings suggest that longer sequences facilitated the segregation of the A and B sounds (i.e. there was evidence of build-up). Chatterjee et al. (2006) observed evidence of build-up in one CI listener, who was instructed to qualitatively report whether a given sequence of sounds was integrated or segregated. In the present study, a detection task was used to assess stream segregation *objectively* (Cooper and Roberts, 2009; Hong and Turner, 2009; Micheyl and Oxenham, 2010a; Roberts et al., 2002) and 7 CI listeners performed the task. The results from the present study support the observations reported in Chatterjee et al. (2006).

The results presented here are also consistent with the findings from Nie and Nelson (2015), who investigated the effects of amplitude modulation rate and sequence duration on voluntary stream segregation in NH listeners. Nie and Nelson (2015) used modulated bandpass noise bursts to simulate the degraded spectral cues present in electric hearing. With a similar task and using similar sequence durations, both studies found an interaction between the temporal cue (amplitude modulation or pulse rate difference) and the sequence duration, suggesting that a similar build-up process might be experienced by CI listeners and NH listeners. Nie and Nelson (2015) found that spectral cues (i.e. a difference between the center frequencies of the noise bands) elicited a build-up both in the presence and in the absence of temporal cues. However, temporal cues elicited a build-up when combined with moderate spectral differences, but not in the absence of spectral cues, suggesting that temporal cues could be a weaker or secondary cue for the segregation of sounds. In the present study, temporal cues elicited a build-up even in the absence of place cues.

Shorter Δt values were needed in the control condition (no distractor) to avoid ceiling effects. This reflects the difficulty experienced by the CI listeners in performing the task in the presence of a distractor stream, even when a large Δrate and a long sequence duration were used. Thus, even though CI listeners seem to be able to achieve a segregated percept and exhibit a similar build-up process as NH listeners, they are not able to completely ignore a competing stream, which may reflect a slower build-up in CI listeners than in NH listeners.

3.4.3 Contribution of temporal regularity differences to stream segregation

In the present study, a temporally irregular distractor stream was used to ensure that temporal judgements between the A and the B sounds would be an unreliable cue to perform the task. Temporally irregular patterns are more likely to be segregated than predictive and temporally regular patterns (for a review, see Bendixen, 2014). Even though the temporal irregularity of the distractor stream cannot account for the increase of the d' scores associated with larger Δrate values, it is possible that the CI listeners made use of both Δrate and regularity differences to segregate the streams. Nie et al. (2014) investigated the role of spectral separation for stream segregation in NH listeners with a paradigm similar to the one used in the present study. Their results suggested that NH listeners could segregate the sounds when the only available cue was the temporal regularity of one stream vs the temporal irregularity of the other. This condition is similar to the *no-difference* condition from the present study. Nevertheless, in the present study, the results from the *no-difference* condition suggest that CI listeners were not able to segregate the streams when the A and B streams were presented through the same electrode and at the same pulse rate. Thus, even though temporal regularity differences between the streams could contribute to their segregation, this cue was not sufficiently salient for it to elicit a segregated percept in the absence of pulse rate differences.

3.4.4 Place vs. temporal cues in stream segregation: the role of pitch differences

The study from Paredes-Gallardo et al. (2018a) (i.e. chapter 2) and the present study employed the same paradigm to assess the role of electrode separation and the role of pulse rate differences in voluntary stream segregation. In electric hearing, both electrode and stimulation pulse rate contribute to the perception of pitch height (Lamping et al., 2018). If stream segregation is correlated with the overall perceptual difference between the sounds, the perceptual Δpitch between the target and the distractor stream may account for the results obtained in the two studies. To test this, the Δrate and the electrode separation values from the present study and from Paredes-Gallardo et al. (2018a) (i.e. chapter 2) were converted to pitch height differences between the target and the distractor streams. Data from a verbal attribute magnitude estimate experiment (Lamping et al., 2018) were used to map specific single electrode stimuli to a perceptual pitch height scale (see supplementary material for more details).

Figure 3.5 shows the d' scores for the long sequence as a function of Δpitch between the target and the distractor streams. On the basis of the magnitude estimation experiment for pitch height, the Δpitch values were normalized such that a Δpitch value of 100% corresponded to the perceived Δpitch between electrodes 11 and 22 at a pulse rate of 900 pps. The data from the CI listeners and predictions from the IO model are indicated by the blue and green boxes, respectively. The pitch differences elicited by varying the pulse rate are shown with a lighter color than pitch differences elicited by changing the stimulation electrode. The solid and dashed lines represent the estimates from the statistical model and its 95% confidence intervals, respectively. The cue used to elicit the pitch differences (electrode vs pulse rate differences) was found to be a non-significant factor [$F(1, 107.31) = 1.216$, $p = 0.273$]. No significant interaction was found between the cue and Δpitch [$F(1, 105.90) = 0.295$, $p = 0.588$], the cue and data type [$F(1, 105.40) = 1.101$, $p = 0.296$] or the cue, Δpitch and data type [$F(1, 98.26) = 0.004$, $p = 0.950$]. Only the Δpitch [$F(1, 6.51) = 18.166$, $p = 0.004$], data type (listeners' data vs IO model predictions) [$F(1, 11.08) = 5.236$, $p = 0.043$] and their interaction [$F(1, 101.85) = 38.612$, $p < 0.001$] were found to be significant effects in the model.

The d' scores from the listeners increased for larger Δpitch values. Moreover, the cue used to elicit the pitch difference was revealed to be a non-significant factor in the statistical model. This suggests that CI listeners can use both place and temporal cues to segregate the streams as long as the perceptual pitch difference between the streams is larger than the fission boundary (i.e. about 20% of the pitch difference between electrodes 11 and 22), supporting the hypothesis proposed by Moore and Gockel (2002, 2012). These findings suggest that the combination of cues may improve stream segregation for CI listeners, provided that a larger overall perceptual difference is elicited between the sounds.

Six of the listeners from the present study had previously participated in the study presented in chapter 2. Thus, a learning effect might have affected the d' scores obtained by the listeners in the present study. Nevertheless, the lack of a

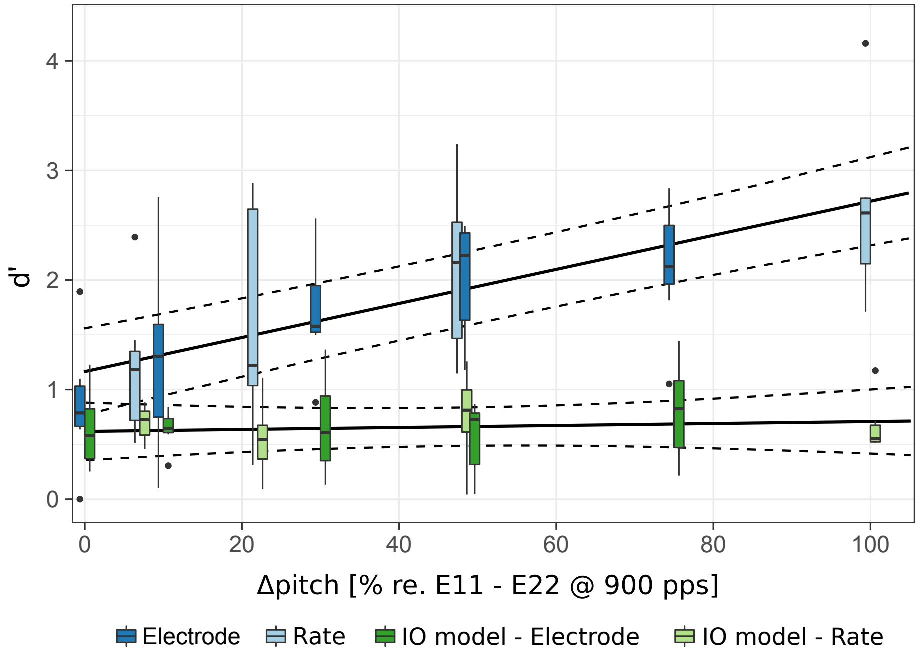


Figure 3.5: Sensitivity (d') scores to the delayed B-sound for each Δpitch between the target and the distractor streams. A Δpitch of 100% corresponds to the pitch difference experienced between electrodes 11 and 22 when stimulated with a pulse rate of 900 pps. The boxes illustrate data from the listeners (in blue) and the corresponding IO model predictions (in green). Dark colors represent the data from chapter 2 (i.e. electrode separation) and light colors represent the data from the present study (i.e. pulse rate differences). The solid lines represent the lsmeans estimate from the statistical model. Its 95% confidence interval is indicated with dashed lines.

significant effect of the cue used to elicit the pitch difference in the combined data from both studies implies that there was not a systematic change in the d' scores from the two studies.

3.5 Summary and conclusion

The present study assessed the effect of temporal cues on voluntary stream segregation in CI listeners. The results suggested that CI listeners can make use of temporal cues to segregate sounds when attention is directed towards segregation. Moreover, a build-up process similar to that reported in NH listeners was observed. The similarity between the trends observed in the present study for CI listeners and those reported for NH listeners suggest a common underlying mechanism for stream segregation in both groups. Furthermore, differences in the perceived pitch height accounted for the results from the present study (temporal cues) as well as from the study presented in chapter 2 (place cues).

This suggests that stream segregation is directly related to the salience of the perceptual difference between the sounds. Thus, the combination of cues may improve stream segregation in CI listeners.

Acknowledgments

We would like to thank all volunteers who participated in this study. We thank the two anonymous reviewers for the helpful and constructive comments on an earlier version of the article, Wiebke Lamping for sharing the data from the verbal attribute magnitude estimation experiment (and for many fruitful discussions) Per B. Brockhoff, Alexandra Kuznetsova and Anders Stockmarr for their help and guidance with the statistical analysis of the data, and our colleagues from the Hearing Systems Group for valuable comments and stimulating discussions. This work was supported by the Oticon Centre of Excellence for Hearing and Speech Sciences (CHeSS) and the Carlsberg Foundation. The research equipment was provided by Cochlear Ltd.

Supplementary material: Perceptual mapping of place and temporal cues

The Δrate and the electrode separation values from the present study and from Paredes-Gallardo et al. (2018a) (i.e. chapter 2) were converted to pitch height differences between the target and the distractor streams (Δpitch). Data from a verbal attribute magnitude estimation experiment (Lamping et al., 2018) were used to map specific single electrode stimuli to a perceptual pitch height scale. Lamping et al. (2018) collected responses from five CI listeners who were instructed to rate the pitch height of single electrode stimuli on a scale from 0 to 100. A rating of 100 reflected full agreement with the verbal attribute high and a rating of 0, no agreement. A combination of four electrodes (i.e. 10, 14, 18, 22) and five pulse rates (i.e. 80, 150, 300, 600 and 1200 pps) were tested. A mixed-effects quadratic model was fitted to the median of the individual ratings over eight repetitions using the statistical software `R` (`lme4` and `lmerTest` libraries: Bates et al. (2014) and Kuznetsova et al. (2017)). Both stimulation electrode, pulse rate (log transformed) and their interaction were treated as fixed effects terms. Listener-related effects were treated as random effects with random intercepts and slopes. Pitch height ratings were defined as the lsmean estimates

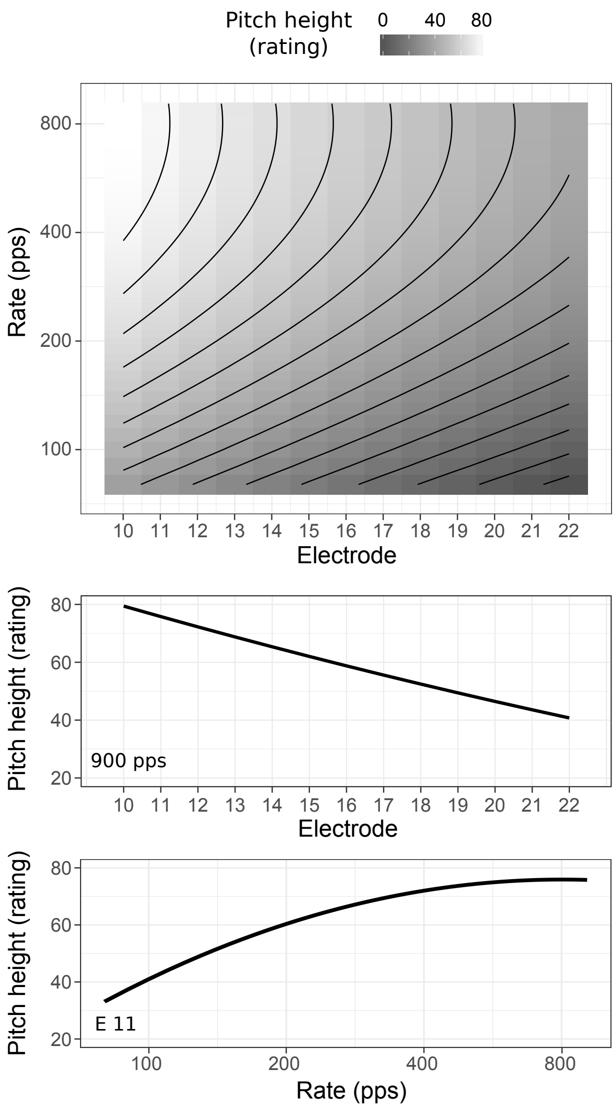


Figure 3.6: Upper panel: predictions of the pitch height ratings for different electrode and pulse rate combinations as obtained from the mixed-effects model fitted to the data from Lamping et al. (2018). Darker colors represent low ratings while brighter colors represent high ratings. The solid lines represent equal-rating contours. Middle panel: predictions of the pitch height ratings for different electrodes at a pulse rate of 900 pps. Bottom panel: predictions of the pitch height ratings for different pulse rates at electrode 11.

of the model (`lsmeans` library: Lenth (2016)) for each target and distractor stimuli. The Δpitch between each target and distractor sound was then calculated. Δpitch values were normalized such that a Δpitch of 100 would correspond to the pitch difference between electrodes 11 and 22 when stimulated at a pulse rate of 900 pps.

Figure 3.6 shows the model predictions of pitch height ratings as a function of electrode and pulse rate (upper panel). High ratings are shown in white while low ratings are shown in dark gray. Equal-rating contours are indicated by the solid black lines. The middle and bottom panels show pitch height ratings for a fixed pulse rate and for a fixed electrode, respectively. The ratings of the pitch height decrease linearly as a function of stimulation electrode (middle panel). Conversely, pulse rate and pitch height exhibit a nonlinear relation (bottom panel) consistent with other studies (e.g. Landsberger et al., 2016). Pitch height ratings increase up to a pulse rate of 300/400 pps and saturate for higher pulse rates.

4

Electrode separation as a cue for auditory stream segregation: Evidence from behavioral measures and event-related potentials[e]

Abstract

The role of electrode separation in sequential stream segregation for cochlear implant (CI) listeners was explored using a deviant detection task. Twelve CI listeners were instructed to attend to a series of target sounds in the presence of interleaved distractor sounds. A deviant was randomly introduced in the target stream either at the beginning, middle or end of each trial. The listeners were asked to detect sequences that contained a deviant and to report its location within the trial. The perceptual segregation of the streams should, therefore, improve deviant detection performance. The electrode range for the distractor sounds was varied, resulting in different amounts of overlap between the target and the distractor streams. For the largest electrode separation condition, event-related potentials (ERPs) were recorded under active and passive listening conditions. The listeners were asked to perform the behavioral task for the active listening condition and encouraged to watch a muted movie for the passive listening condition. Deviant detection performance improved by increasing electrode separation between the streams, suggesting that larger electrode differences facilitate the segregation of the streams. Deviant detection performance was best for deviants happening late in the sequence, indicating that a segregated percept builds up over time. The analysis of the ERP

[e] This chapter is based on Paredes-Gallardo et al. (under review for Frontiers in Neuroscience).

waveforms revealed that auditory selective attention modulates the ERP responses in CI listeners. Specifically, the responses to the target stream were, overall, larger in the active relative to the passive listening condition. Conversely, the ERP responses to the distractor stream were not affected by selective attention. However, no significant correlation was observed between the behavioral performance and the amount of attentional modulation. Overall, the findings from the present study suggest that CI listeners can use electrode separation to perceptually group sequential sounds. Moreover, the listeners can selectively attend to the stream of interest, as reflected by the attentional modulation of the ERPs at the group level.

4.1 Introduction

Many daily listening scenarios involve multiple sound sources. Thus, to selectively listen to a single person's voice among many, or to a melody in a complex musical arrangement, the listener needs to parse the sounds in the complex auditory scene and group them into meaningful auditory objects or streams (e.g. McDermott, 2009). This process is known as auditory scene analysis (Bregman, 1990). Hearing impairment may affect the process of object formation and thus, hearing-impaired (HI) listeners generally perform worse than normal-hearing (NH) listeners in complex listening scenarios (e.g. Mackersie et al., 2001; Oxenham, 2008). This is the case even when hearing aids or cochlear implants (CIs) are used to make the signals audible (e.g. Nelson et al., 2003). Most current CIs convey spectral information through place cues, whereby different frequency bands of the acoustic signal stimulate different electrodes at a given pulse rate. It is unclear to what extent CI listeners can use place cues in the process of object formation. Thus, a better understanding of the role of place cues in the process of object formation would be beneficial to overcome the challenges that CI listeners experience in complex listening scenarios.

Several studies have investigated obligatory stream segregation in CI listeners (e.g. Cooper and Roberts, 2009; Duran et al., 2012; Tejani et al., 2017). These studies varied either the spatial separation between the stimulating electrodes (i.e. place cues) or the pulse rate of stimulation at a fixed electrode (i.e. temporal cues). The manipulation of these parameters is known to elicit perceptual differences in CI listeners (e.g. Eddington et al., 1978; Landsberger et al., 2016).

The results from Tejani et al. (2017) were consistent with those from studies in NH listeners, i.e. the probability of experiencing a two-stream percept increased by increasing the electrode separation. Similarly, the results from Duran et al. (2012) suggest that CI listeners experience obligatory stream segregation when only temporal cues are provided. However, Cooper and Roberts (2009) did not observe a build-up effect, suggesting that not all elements of obligatory stream segregation may be experienced by CI listeners.

Cooper and Roberts (2009) also assessed voluntary stream segregation abilities of CI listeners using a melody discrimination task. The listeners were asked to identify a pattern of sequentially activated electrodes (melody) in the presence of interleaved, random distractor sounds. Their results showed that CI listeners were not able to segregate the melody from the distractor sounds, regardless of the electrode separation between the streams. Conversely, other studies suggest that CI listeners can use either electrode separation, pulse rate differences or amplitude modulation (AM) rate differences to voluntarily segregate sequences composed of two repeating and alternating sounds (Hong and Turner, 2009; Paredes-Gallardo et al., 2018a,b). Moreover, a build-up effect has been reported during voluntary stream segregation when using either place or temporal cues (Paredes-Gallardo et al., 2018a,b).

There are several differences between the study of Cooper and Roberts (2009) and those of Hong and Turner (2009) and Paredes-Gallardo et al. (2018a,b). The sequences of sounds were shorter in the study of Cooper and Roberts (2009) than those used in the studies of Hong and Turner (2009) and Paredes-Gallardo et al. (2018a,b). Thus, it is possible that the poor performance reported in the study of Cooper and Roberts (2009) reflects the long time that CI listeners need to build up a segregated percept, even when the attention of the listener is directed towards segregation (Paredes-Gallardo et al., 2018a,b). Another difference is that Cooper and Roberts (2009) used streams composed of different sounds, resulting in a more complex task than those employed in the studies of Hong and Turner (2009) and Paredes-Gallardo et al. (2018a,b). Finally, the stimuli used by Hong and Turner (2009) and Paredes-Gallardo et al. (2018a,b) might have facilitated segregation due to the inclusion of rhythmic cues (for a review, see Bendixen, 2014). Thus, it is unclear whether CI listeners are able to segregate streams composed of different sounds in the absence of rhythmic cues.

It has been suggested that selective attention operates as a form of sensory gain control, modulating the neural representations of signals in the auditory cortex. Specifically, selective attention has been shown to enhance the event-related potentials (ERPs) evoked by attended sounds and to suppress those evoked by ignored sounds (e.g. Choi et al., 2013; Hillyard et al., 1998; Hillyard et al., 1973; Picton and Hillyard, 1974). Moreover, the amount of attentional modulation of the ERPs has been shown to correlate with the listener's ability to perform an auditory selective-attention task (Choi et al., 2014; Dai and Shinn-Cunningham, 2016; Dai et al., 2018), suggesting a strong link to perception.

The present study investigated 1] whether CI listeners can use electrode separation as a cue to segregate streams composed of different sounds, 2] whether a two-stream percept builds-up over time, 3] whether selective auditory attention modulates the amplitude of the ERPs and 4] whether such attentional modulation of the ERP reflects individual stream segregation abilities (i.e. whether the attentional modulation of the ERPs can be used as an objective tool to assess stream segregation abilities). Behavioral detection performance was measured in a paradigm where the listeners were required to attend to a series of sounds in the presence of interleaved distractor sounds. A deviant was randomly introduced in the target stream either at the beginning, middle or end of each trial. The listeners were asked to detect sequences that contained a deviant and to report its location within the trial. As in the task described by Cooper and Roberts (2009), the perceptual segregation of the streams should improve performance in the deviant detection task. It was hypothesized that if CI listeners can use electrode separation as a cue to segregate the streams, performance in the deviant detection task would improve with increasing electrode separation between the target and the distractor streams. If CI listeners need time to build up a segregated percept, detection performance should be highest for deviants presented late in the trial. Furthermore, ERPs to the same stimuli were recorded while the listeners performed the behavioral task (active listening) and while they watched a muted movie (passive listening). It was hypothesized that if CI listeners can segregate the streams, then the ERPs evoked by the target stream should be enhanced in the active listening condition compared to the passive listening condition. Conversely, ERPs evoked by the distractor stream should be suppressed in the active listening condition with respect to the passive listening condition.

4.2 Methods

The experiments took place in a sound-attenuating and electrically shielded booth at the Bionics Institute of Australia and at the Technical University of Denmark. The experiments were conducted in three sessions, each lasting 2h including short breaks. The first session comprised categorical loudness scaling and loudness matching of the different stimuli, a pitch ranking task, a test run of the detection task in the absence of the distractor stream and a 15 – 20 min training on the segregation task. The behavioral experiment and the recording of the ERPs took place in the second and third sessions, respectively.

4.2.1 Listeners

Twelve CI listeners participated in this study. The listeners were aged between 20 and 82 years (mean: 61.3 years, SD: 22.2 years; see table 4.1), had no residual hearing in the implanted ear and were users of the Cochlear Ltd. (Sydney, Australia) implant. For the bimodal listeners, the contralateral ear was unaided and blocked with an earplug during the experiments. All listeners performed the behavioral task. Listener CI-10 did not participate in the ERP recording session. All listeners provided written informed consent prior to the study and all experiments were approved by the Human Research Ethics Committee of the Royal Victorian Eye and Ear Hospital (reference 14.1180H) and the Science-Ethics Committee for the Capital Region of Denmark (reference H-16036391).

4.2.2 Task description

The listeners were asked to perform a detection task, illustrated in figure 4.1. The target stream consisted on a pattern of sounds, presented on electrodes 9, 11 and 13 (i.e. a triplet). On each trial, three triplets were presented consecutively in the presence of interleaved, random distractor sounds which the listeners were asked to ignore. Each triplet began with a target sound and ended with a distractor sound. Both the target and the distractor streams extended over a range of five electrodes. A deviant triplet was randomly introduced in 75% of the trials by reversing the electrode sequence in one of the three triplets (i.e. electrodes 13, 11 and 9). In the remaining 25% of the trials, no deviant was presented, i.e. the three triplets were identical (see figure 4.1). The listeners were asked to detect sequences that contained a deviant and to report its location

Table 4.1: Information about the participants regarding age, onset of deafness, implanted ear, number of years of experience and modality of rehabilitation.

ID	Age range	Onset of deafness	Implant model (ear)	Years of experience	Modality
CI 1	30-49	Postlingual	CI24RE (right)	4	Bilateral
CI 2	> 70	Postlingual	CI522 (left)	1	Bimodal
CI 3	< 30	Postlingual	CI522 (left)	1	Unilateral
CI 4	50-69	Postlingual	CI24RE (left)	4	Bimodal
CI 5	> 70	Postlingual	CI512 (left)	1	Bimodal
CI 6	> 70	Postlingual	CI512 (right)	2	Bilateral
CI 7	50-69	Postlingual	CI24RE (left)	3	Bilateral
CI 8	> 70	Postlingual	CI24RE (right)	8	Bimodal
CI 9	> 70	Perilingual	CI24RE (right)	3	Bilateral
CI 10	< 30	Prelingual	CI24RE (left)	7	Unilateral
CI 11	50-69	Perilingual	CI24RE (left)	8	Bilateral
CI 12	< 30	Postlingual	CI24R (left)	14	Bilateral

within the trial in a one-interval, four-alternative forced-choice paradigm.

Each sound consisted of a 50 ms burst of biphasic pulses presented at a given electrode. Each biphasic pulse had a phase width of 25 μs and an inter-phase gap of 8 μs. The pulse rate was fixed at 900 pps. The inter-stimulus interval (ISI) was 440 ms between two consecutive target or distractor sounds, and 220 ms between two consecutive target-distractor sounds (see figure 4.1). The stimuli were presented in monopolar mode through the Nucleus Implant Communicator research interface (NIC v3, Cochlear Ltd, Sydney) and a research speech processor (L34) provided by Cochlear Ltd.

4.2.3 Loudness balancing

Previous studies have suggested that loudness could be an effective cue for the segregation of sounds for CI listeners (e.g. Cooper and Roberts, 2009; Marozeau et al., 2013). To ensure that the listeners did not rely on loudness cues to segregate the sounds, the stimuli of the present study were loudness-balanced. A total of 16 electrodes (from electrode 5 to electrode 20) were used in the present study. Categorical loudness scaling was used to find the most comfortable level (MCL) for six electrodes (i.e. electrodes 5, 8, 11, 14, 17 and 20) using an 11-step attribute scale, as described in chapter 2. The MCL for the remaining

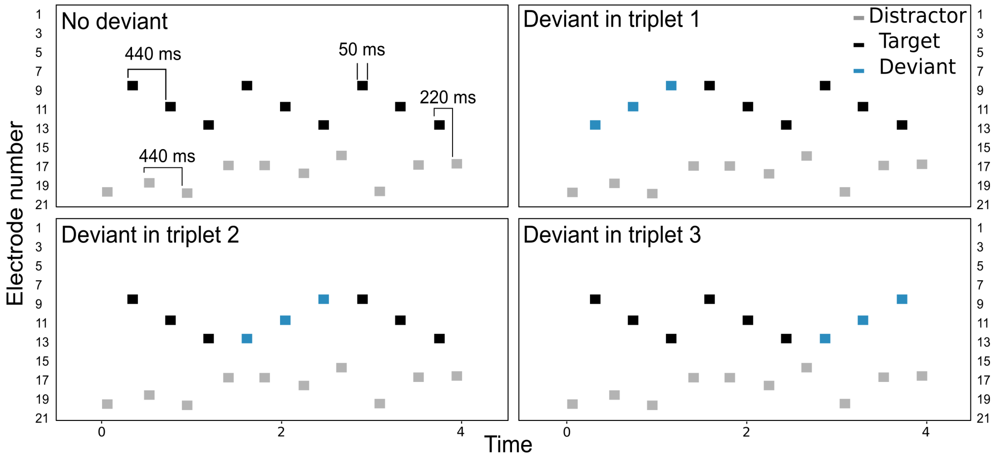


Figure 4.1: Schematic representation of the electrodogram for each of the four deviant conditions. Black and gray markers represent the target and the distractor sounds, respectively. The deviant triplet is shown with blue markers.

electrodes (i.e. electrodes 6, 7, 9, 10, 12, 13, 15, 16, 18 and 19) was obtained by linear interpolation. All electrodes were then loudness matched to a reference electrode (electrode 11) by the listeners, using a simple user interface. The interface allowed the increase and the decrease of the test-sound intensity in steps of 0.15, 0.3, or 0.45 dB.

4.2.4 Inclusion criteria

Most CI listeners report a monotonic relation between the place of stimulation and the corresponding pitch percept (e.g. Eddington et al., 1978; Shannon, 1983; Tong et al., 1980; Townshend et al., 1987). However, several previous studies reported instances where the pitch percept did not follow a monotonic function (e.g. Collins et al., 1997; Nelson et al., 1995). In the present study, local pitch reversals could hinder the performance in the detection task. Thus, a pitch ranking experiment was conducted with eight odd-numbered electrodes (between electrodes 5 and 20) using the midpoint comparison procedure (Long et al., 2005; Macherey and Carlyon, 2010). All twelve listeners exhibited monotonic pitch ranks. Furthermore, to ensure that all listeners were able to perform the detection task, a test run with 20 presentations of each of the four different conditions (see figure 4.1) was performed in the absence of the distractor stream. A minimum average performance of 90% correct was achieved by all listeners.

4.2.5 Behavioral experiment

Stimuli and conditions

In the behavioral experiment, four electrode separation conditions between the target and the distractor streams were tested. The four conditions are represented in figure 4.1. The target stream was always presented on electrodes 9, 11 and 13, whereas the electrode range of the distractor stream was varied. In the *no overlap* condition, the distractor stream was presented through electrodes 16 to 20, resulting in a separation of three or more electrodes between the streams. In the *apical overlap* condition, the distractor stream was presented through electrodes 13 to 17, resulting in an overlap with the most apical electrode of the target stream. In the *basal overlap* condition, the distractor stream was presented through electrodes 5 to 9, such that there was an overlap with the most basal electrode of the target stream. Finally, in the *full overlap* condition, the electrode range of both target and distractor streams was identical (i.e. electrodes 9 to 13).

In each trial, a random number of one to four distractor sounds was played before and after the target stream (i.e. inducer sounds). Thus, the listeners did not have a priori knowledge about the starting point of the target stream. This was done to encourage the listeners to attend to the full duration of the trial instead of listening for a specific time point. The duration of each trial ranged between 4 s and 6.65 s.

Procedure

Behavioral responses during both the initial training and the data collection were recorded using a custom-made user interface in Python. Four response buttons were used to record the listener's response 200 ms after each trial. The duration of the inter-trial interval was randomized between 1.5 and 2.5 s. Feedback was provided after each trial.

Prior to the behavioral experiment, the listeners underwent 15 to 20 min of training in the stream segregation task. The training began with the detection task in the absence of the distractor stream. Once the listeners were familiarized with the sequences, the distractor stream was introduced at a soft, but audible, level (*no overlap* condition). The level of the distractor stream was increased in steps of 0.45 dB every third sequence until both streams where played at the listener's MCL. The training procedure was repeated with the three remaining

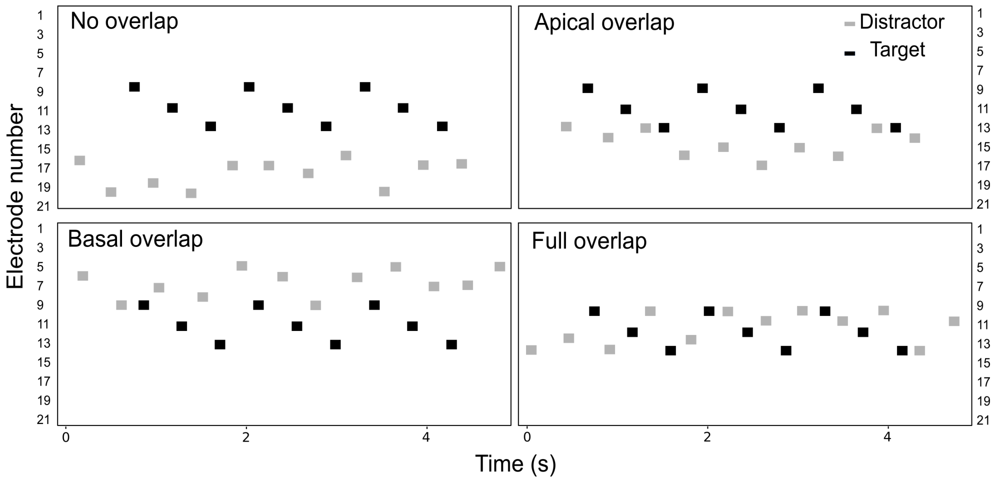


Figure 4.2: Schematic representation of the electrodogram for each of the four electrode separation conditions. Black and gray markers represent the target and the distractor sounds, respectively. On each trial, the distractor stream started before the target stream and ended after the target stream, with a random number of one to four sounds.

distractor sets, i.e. apical, basal and full overlap.

In the behavioral experiment, a total of 20 trials were presented for each electrode separation and each of the four deviant conditions. The resulting 320 trials were divided into eight blocks. A block consisted of 10 trials of each of the four deviant conditions for a given electrode separation condition. The order of the blocks was randomized.

Data analysis

The sensitivity measure (d') was calculated using equation (4.1) for each of the three deviant triplet locations (i), where z represents the z-transformation, N_{H_i} and N_{FA_i} the number of hits and false alarms, respectively, and H and FA the maximum number of hits and false alarms (20 and 60, respectively). The log-linear rule was used to avoid undefined extremes when the hit or the false alarm rates take the values of zero or one (Hautus, 1995; Verde et al., 2006).

$$d'_i = z\left(\frac{N_{H_i}+0.5}{H+1}\right) - z\left(\frac{N_{FA_i}+0.5}{FA+1}\right) \quad (4.1)$$

Statistical inference was performed by fitting a mixed-effects linear model to the d' scores. The experimental variables and their interactions were treated as fixed effects whereas listener-related effects were treated as random effects with random intercepts and slopes. The model was implemented in R using the lme4 library (Bates et al., 2014) and the model selection was carried out with the lmerTest library (Kuznetsova et al., 2017) following the backwards selection approach based on step-wise deletion of model terms with high p-values (Kuznetsova et al., 2015). The p-values for the fixed effects were calculated from F-tests based on Sattethwaite's approximation of denominator degrees of freedom and the p-values for the random effects were calculated based on likelihood ratio tests (Kuznetsova et al., 2015). The post-hoc analysis was performed through contrasts of least-square means using the lsmeans library (Lenth, 2016). The p-values were corrected for multiple comparisons using the Tukey method.

4.2.6 Recording of event-related potentials

Stimuli and conditions

The *no overlap* condition was chosen for the recording of the ERPs. Thus, the target stream was presented at electrodes 9, 11 and 13 and the distractor stream comprised electrodes 16 to 20. It has been suggested that the first sound of a sequence may draw attention exogenously (e.g. Choi et al., 2014; Dai et al., 2018). In order to ensure that the listeners deployed top-down attention to the target stream, on each trial, a single distractor sound was played before the first triplet. Thus, in the present study, the target stream was always the lagging stream.

A 50 ms burst of pulses on electrode 11, followed by a 750 ms silence was played before each trial. This burst was not relevant for the behavioral task, and the listeners were not given specific listening instructions (whether to attend or ignore it). The burst was included to normalize the N1 amplitude for the remaining sounds across listeners (e.g. Choi et al., 2014). It was hypothesized that the N1 response elicited by this burst would reflect individual differences in the N1 amplitude but would not be affected by attention. However, the N1 responses to this burst were affected by attention, and this effect was variable across listeners. Thus, the responses to this pre-trial burst were not used for the normalization of the individual N1 amplitudes.

Procedure

Two attention conditions were tested in the ERP recording session: an active listening condition and a passive listening condition. During the active listening condition, the scalp electroencephalogram (EEG) was recorded while the listeners performed the detection task. In the passive listening condition, the EEG signal was recorded in response to the same sounds while the listeners watched a muted movie with captions. Overall, a total of 55 trials were recorded for each attention (active/passive) and deviant condition. For the active listening condition, the 220 trials were divided into two blocks of 110 trials each (15 min). For the passive listening condition, the 220 trials were recorded in a single block (30 min).

EEG data acquisition and analysis

The EEG data were recorded using the Biosemi ActiveTwo™system at a sampling rate of 8192 Hz. The hardware anti-aliasing filter bandwidth follows a 5th order sinc response, with the -3 dB point located at 1600 Hz. 68 electrodes were used for the recording: 64 electrodes mounted on an elastic headcap according to the international 10-20 electrode configuration, two electrodes at the left and right mastoids, one electrode near the outer canthus of the eye and one electrode below the eye contralateral to the CI. The electrodes directly over the coil were not used in the recording and all electrode wires were directed away from the coil to minimize radio frequency artefact. The offsets of the recording electrodes were kept below 20 mV in all recordings.

The data were processed using the Fieldtrip toolbox (Oostenveld et al., 2011) and customized Matlab scripts. The continuous EEG data were re-referenced to the average mastoids, highpass-filtered at 1 Hz (FIR with zero-phase lag, 1 Hz transition bandwidth) and lowpass-filtered at 100 Hz (FIR with zero-phase lag, 1 Hz transition bandwidth). The data were downsampled to 256 Hz and epoched from -0.5 to 5.5 s relative to the sound onset. The total duration of the epoch was used for baseline subtraction. Epochs containing unique, non-stereotyped artifacts were manually rejected. Infomax independent component analysis (ICA) was then applied to the remaining epochs. Equivalent current dipole modelling was computed for all independent components (ICs) on each condition. ICs representing eye blinks and saccadic eye movement were manually identified based on their scalp topography, waveform and power-spectrum. Components representing the radio frequency artifact from the implant were automatically identified with a custom implementation of the procedure described in Viola et al. (2012) (see supplementary material for more details). Artefactual components were removed from all datasets and data were back-projected to the sensor space. After artifact correction, the time interval between -205 and -5 ms was used for baseline subtraction. Epochs were lowpass-filtered at 20 Hz (FIR, zero-phase lag, 1 Hz transition bandwidth and 1 s zero-padding both before and after the epoch).

Only the correctly answered trials were processed. Since there were not enough correct trials to analyze the data for each of the four deviant conditions, the epochs were grouped in *early* and *late* deviant conditions. The *early* deviant group contained the epochs where either the first or the second triplets were a deviant. The *late* deviant group contained the epochs where the deviant was in the third triplet or absent. A minimum of 62 correct trials was available for each listener, deviant condition (*early* vs. *late*) and attention condition (active vs. passive). Thus, the first 62 correctly answered trials for each listener, deviant condition and attention condition were analyzed. For each condition, listener and electrode, the amplitude of the N1 ERP component was calculated as the local minimum in the time window from 70 to 170 ms after each sound onset. For each listener, the across-electrode N1 amplitude was calculated by averaging the amplitudes from nine front-central electrodes (Fz, AFz, FCz, F1, F2, FC1, FC2, AF3, AF4). This average measure will be referred to as N1 amplitude.

The attentional modulation of the ERPs was quantified for each listener as the difference in N1 amplitude between the active and passive listening conditions for each sound. Thus, a negative value indicates a larger N1 response in the active than in the passive listening condition. These values were averaged across all target sounds to obtain a single estimate per listener. The single value was used to compute the Kendall rank correlations between the behavioral performance and the attentional modulation of the ERPs.

A mixed-effects linear model was used for the statistical analysis. N1 amplitude differences between the active and the passive listening conditions were modeled following the approach described in section 4.2.5.

4.3 Results

4.3.1 Behavioral experiment

The results from the behavioral experiment are shown in figure 4.3. The d' scores are shown for each combination of electrode separation (A) and deviant triplet location (B). Different electrode separation conditions are shown with different colors. Statistically significant differences between conditions are illustrated with letters. Conditions sharing one or more letters are not significantly different. Detailed statistics from the post hoc analysis are provided in the supplementary material.

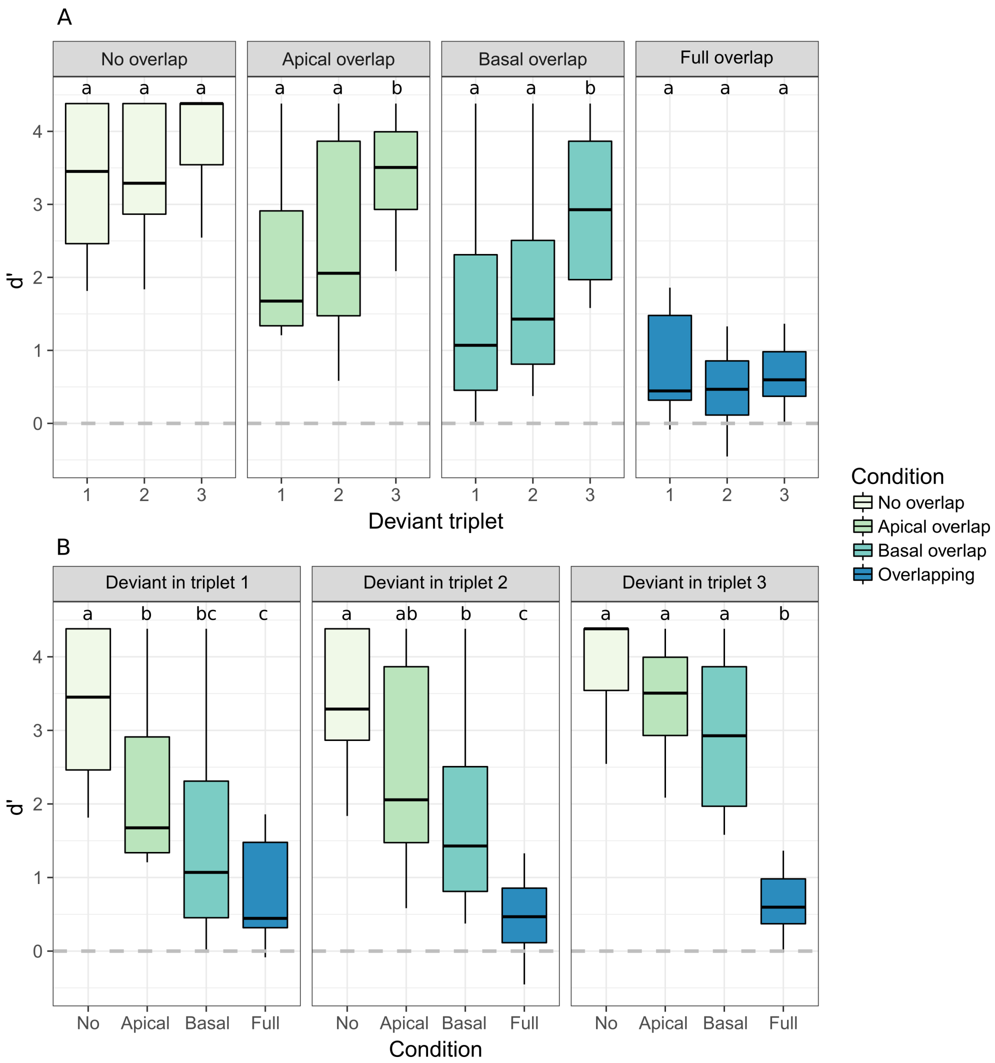


Figure 4.3: Boxplot of the sensitivity scores (d') to the deviant triplet for each electrode separation and deviant triplet location. The color of the boxes represents the electrode separation condition. A] Effect of the deviant triplet for each electrode separation condition. B] Effect of the electrode separation for each deviant triplet location. Results from the statistical contrasts are indicated with lowercase letters. Conditions sharing one or more letter are not significantly different (significance level $\alpha = 0.05$).

Overall, the d' scores increased the later the deviant triplet occurred [$F(2,11.75) = 16.423$, $p < 0.001$] and decreased with increasing electrode overlap between the streams [$F(3,11.39) = 73.484$, $p < 0.001$]. Moreover, a significant interaction was found between the deviant triplet location and the electrode overlap between the streams [$F(6,88.00) = 5.811$, $p < 0.001$], indicating that the effect of the deviant triplet location was not the same for all electrode separation conditions.

The location of deviant triplet did not affect the d' scores for the *no overlap* condition (figure 4.3A). However, the d' scores were at ceiling for this condition, preventing any effect of the deviant triplet location to be observed. Similarly, no effect of the deviant triplet location was observed for the *full overlap* condition, where the d' scores were close to zero. The largest effect of the deviant triplet location was observed for the *apical* and *basal overlap* conditions. In these conditions, significantly larger d' scores were achieved when the deviant triplet happened at the end of the sequence.

When the deviant occurred in the first triplet, the d' scores for the no overlap condition were significantly larger than the ones achieved for any of the other conditions (figure 4.3B). The difference between the *no overlap* condition and the *apical* and *basal overlap* conditions was reduced when the deviant occurred in the second triplet. No significant difference was observed between these three conditions when the deviant occurred in the third triplet. The d' scores were generally lower for the *basal overlap* than for the *apical overlap* condition. However, no significant difference was observed between these two conditions for any of the deviant triplet locations.

4.3.2 Event-related potentials

The grand average waveform across all listeners is shown in figure 4.4 (averaged across all deviant conditions) and in figure 4.5 (for the early and late deviant conditions). Red and blue solid lines represent the active and the passive listening conditions, respectively. Blue and gray shaded areas indicate the N1 response time window for the target and the distractor sounds, respectively. Sharp oscillations before the N1 time window are likely to represent the residual CI artifact, and should not be mistaken for a P1 response. The scalp distribution of the response to one target and one distractor sound is also shown for each listening condition and their difference. The scalp distributions were obtained by averaging the response over the N1 time window for each of the 64 electrodes. Blue and red colors represent negative and positive values, respectively.

N1 responses to the first triplet were, qualitatively, similar for the active

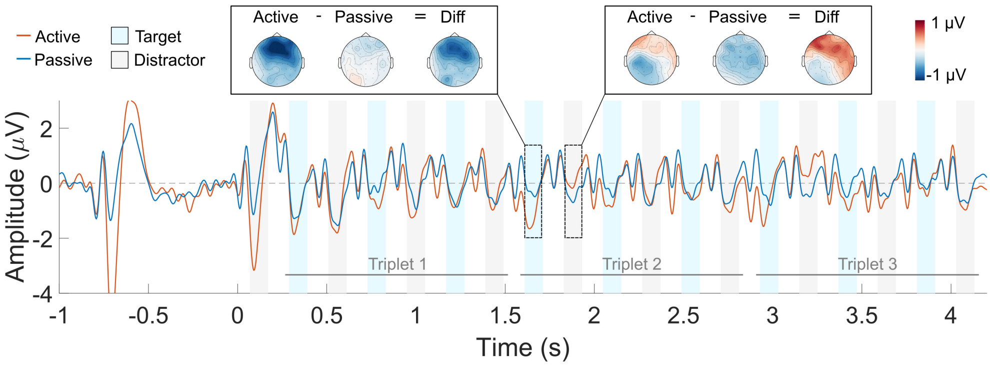


Figure 4.4: Averaged ERP waveform across the four deviant triplet conditions. The active listening condition is shown in red and the passive listening condition in blue. The blue and gray shaded areas indicate the N1 ERP component time window for the target and the distractor sounds, respectively. Each trace represents the average across nine front-central electrodes (Fz, AFz, FCz, F1, F2, FC1, FC2, AF3, AF4). The scalp topography of the response to a target and a distractor sound is shown for each of the listening conditions and their difference.

and the passive listening condition (figure 4.4). This was the case for both the target and the distractor sounds. Conversely, the N1 attentional modulation was different for the target and the distractor sounds in the second and third triplets: selective attention enhanced the N1 responses to the target sounds (i.e. the N1 amplitude is more negative in the active vs. the passive listening condition) and suppressed the N1 responses to the distractor sounds (i.e. the N1 amplitude is more negative in the passive vs. the active listening condition). However, this was only observed for the first two target sounds and the first distractor sound of the second and third triplets. Similar patterns can be seen in figure 4.5, both for the early and for the late deviant conditions. Nevertheless, the effect of attention is largest for the late deviant condition. This is apparent when comparing the topography of the N1 responses to a target and a distractor sound in figure 4.4 and figure 4.5.

The individual N1 attentional modulation for each deviant condition (*early* vs. *late* deviant), sound type (target vs. distractor), triplet number and sound number (first, second or third sound of a triplet) was modeled using a mixed-effects statistical model. The first sound of the sequence was not part of any of the triplets and therefore, was excluded from the analysis. The model revealed a significant main effect of the sound type [$F(1,356) = 20.051$, $p < 0.001$]. Moreover, a significant interaction was found between the triplet number, the deviant condition and the sound type [$F(2,356) = 4.557$, $p = 0.011$] and between the sound number, the deviant condition and the sound type [$F(2,356) = 3.111$,

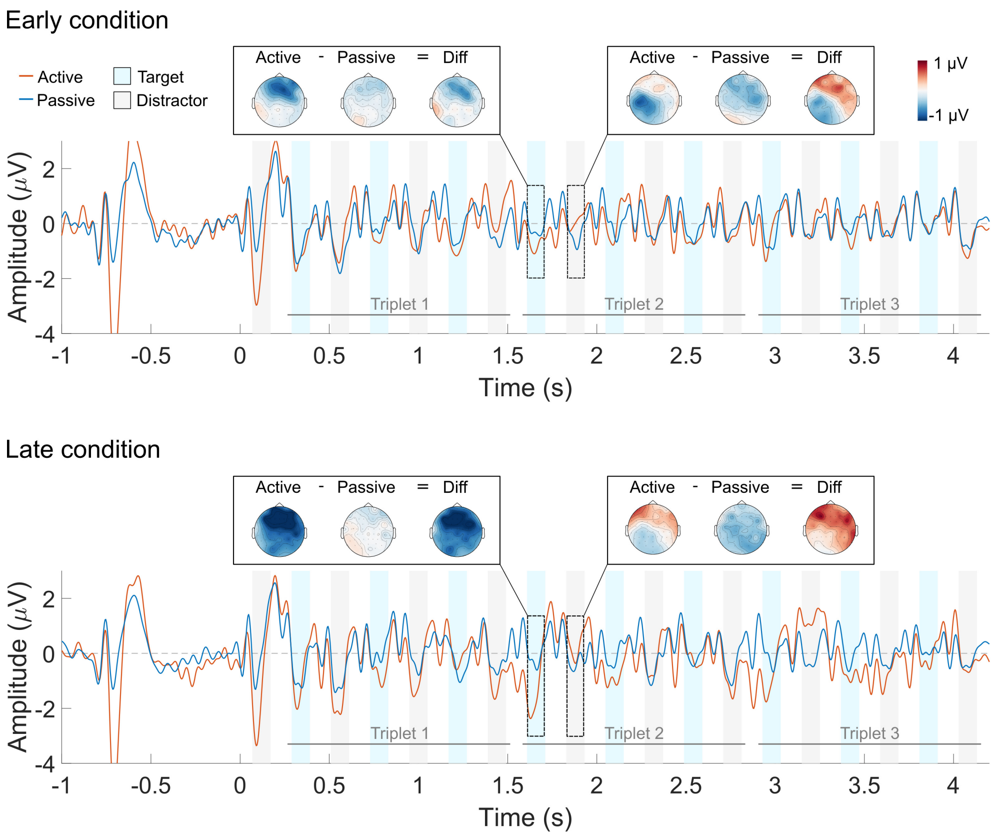


Figure 4.5: Averaged ERP waveform for the *early* (top) and *late* (bottom) deviant conditions. The active listening condition is shown in red and the passive listening condition in blue. The blue and gray shaded areas indicate the N1 ERP component time window for the target and the distractor sounds, respectively. Each trace represents the average across nine front-central electrodes (Fz, AFz, FCz, F1, F2, FC1, FC2, AF3, AF4). The scalp topography of the response to a target and a distractor sound is shown for each of the listening conditions and their difference.

p = 0.046]. No significant interaction was found between the triplet number, the sound number and the deviant condition [F(4,344) = 0.464, p = 0.763] or between the triplet number, the sound number, the sound type and the deviant condition [F(4,340) = 1.049, p = 0.382].

A post hoc analysis revealed that the N1 responses to the target sounds were, on average, enhanced by 0.623 μV in the active vs. the passive listening condition [t(14.9) = 4.336, p = 0.001]. Conversely, the difference in the N1 responses elicited by the distractor sounds was not statistically significant [estimate = 0.075 μV, t(14.9) = 0.524, p = 1]. This was also the case when including the responses to the first sound of the sequence (i.e. a distractor sound) in the analysis.

The significant interactions from the model are illustrated in figure 4.6, where the N1 attentional modulation is shown for each sound type and deviant condition. In figure 4.6A, the N1 attentional modulation is averaged across the three sounds of each triplet, illustrating the interaction between the triplet number, the deviant condition and the sound type. For the *early* deviant condition, no significant modulation of the N1 responses was observed for either the target or the distractor sounds in any of the three triplets. For the *late* deviant condition, a significant N1 enhancement was observed for the first triplet of the distractor stream [t(26.97) = 2.719, p = 0.034] and for the second [t(26.97) = 3.315, p = 0.008] and third [t(26.97) = 2.672, p = 0.038] triplets of the target stream. In figure 4.6B, the N1 attentional modulation is averaged across the three triplets for each sound, illustrating the interaction between the sound number, the deviant condition and the sound type. As in figure 4.6A, no significant modulation of the N1 responses to any of the sounds was observed for the *early* deviant condition. However, a significant N1 enhancement was observed for the first [t(26.97) = 3.763, p = 0.003] and second [t(26.97) = 3.451, p = 0.006] target sounds for the *late* deviant condition.

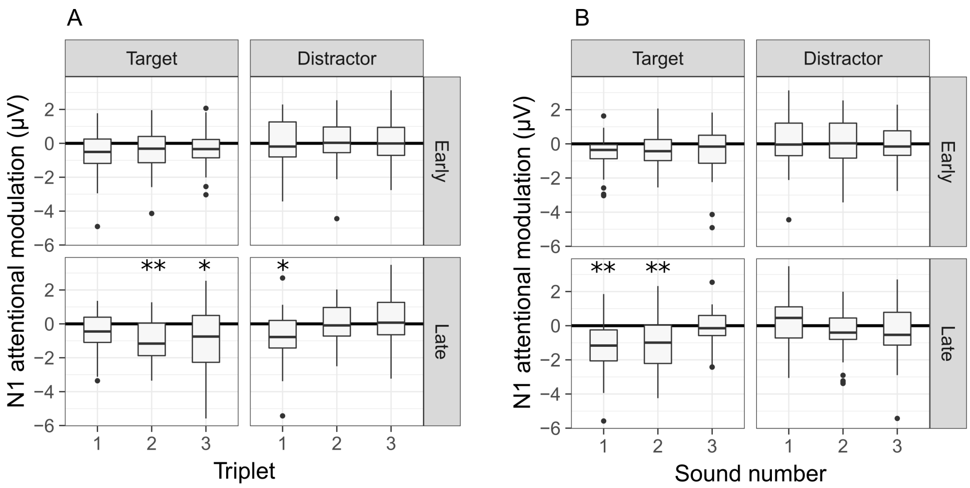


Figure 4.6: N1 attentional modulation of the target and distractor sounds for each deviant condition. A negative value represents an enhanced N1 response in the active condition. A] Averaged N1 attentional modulation across the three sounds of each triplet. B] Averaged N1 attentional modulation across the three triplets for each sound. A statistically significant difference from zero is indicated by one asterisk if 0.05>p>0.01, two asterisks if 0.01>p>0.001 and three asterisks if p<0.001. The p-values were corrected for multiple comparisons using the Bonferroni correction.

The relation between the individual d' scores and the N1 attentional modulation of the target sounds is shown in figure 4.7. No significant correlation was found between the d' scores achieved in the active listening condition and the N1 attentional modulation (figure 4.7A) [τ = -0.037, p = 0.876]. Kendall rank correlation scores were also computed for the N1 attentional modulation and the d' scores from the behavioral session. No significant correlation was found for the no overlap condition (figure 4.7B) [$\tau = 0.112$, p = 0.637], for the apical overlap condition (figure 4.7C) [$\tau = 0.127$, p = 0.648] or for the basal overlap condition (figure 4.7D) [$\tau = 0.273$, p = 0.283].

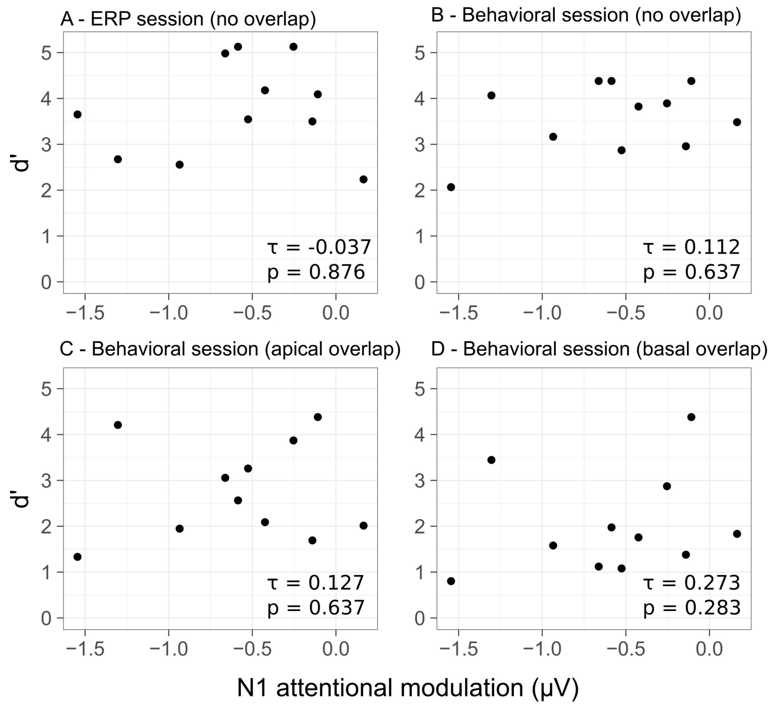


Figure 4.7: Scatter-plots of the d' scores as a function of the average N1 attentional modulation for each listener and condition. Behavioral d' scores from the ERP session are shown in panel A (*no overlap* condition). The d' scores from the behavioral session are shown in panel B for the *no overlap* condition, in panel C for the *apical overlap* condition and in panel D for the *basal overlap* condition. Kendall rank correlation coefficients and p-values are shown in the bottom-right corner of each panel.

4.4 Discussion

4.4.1 The effect of electrode separation on stream segregation

Performance in the detection task was assumed to improve when the target and the distractor streams were perceptually segregated. Overall, the d' scores obtained by the listeners increased with increasing electrode separation between the streams. The d' scores obtained by the listeners were near chance-level for the *full overlap* condition, indicating that the listeners could not segregate the streams in the absence of place cues. Conversely, performance was at ceiling for the *no overlap* condition, suggesting that the listeners were able to use place cues to segregate the streams. These findings are consistent with previous work suggesting that electrode separation facilitates stream segregation for CI listeners (Böckmann-Barthel et al., 2014; Chatterjee et al., 2006; Hong and Turner, 2006; Paredes-Gallardo et al., 2018a,b; Tejani et al., 2017).

Previous studies generally assessed stream segregation using sequences composed of two repeating and alternating sounds. In such paradigms, the listeners need to segregate the target sound from the distractor sound. In contrast, in the present study, each stream was composed of multiple and different sounds, increasing the complexity of the task: The listeners had to integrate different sounds to form a representation of the target and the distractor streams and to maintain these representations segregated over time. Despite the differences between the paradigms, the results of the present study were consistent with those from Paredes-Gallardo et al. (2018a) (i.e. chapter 2), which suggested that most CI listeners can segregate streams separated by three electrodes. In the present study, all listeners could segregate the streams in the *no overlap* condition, where the minimum electrode separation between the streams was three electrodes.

The d' scores obtained by the listeners increased for deviants happening late in the sequence. This suggests that a two stream percept built up over time (for a review, see Moore and Gockel, 2002, 2012). These results are consistent with earlier reports suggesting that CI listeners may experience a build-up when attention is directed towards segregation, i.e., voluntary streaming (Paredes-Gallardo et al., 2018a,b, described in chapters 2 and 3 from this thesis). Moreover, the effect of the deviant triplet location was found to be dependent on the electrode separation between the streams. The largest effect of the deviant triplet location (i.e. build-up) was observed for the *apical overlap* and the *basal overlap* conditions, where the performance was not at ceiling or at chance level. This is consistent with previous reports from studies with NH listeners (e.g. Anstis and Saida, 1985; Bregman, 1978) and CI listeners (Böckmann-Barthel et al., 2014), that reported a build-up for intermediate frequency differences between the sounds. However, Böckmann-Barthel et al. (2014) did not provide specific listening instructions to the participants, who directly reported their perception.

It has been suggested that the results from such paradigms may reflect pitch or electrode discrimination instead of stream segregation (Chatterjee et al., 2006; Cooper and Roberts, 2007). This uncertainty was avoided in the present study by using a detection task which was facilitated by the segregation of the sounds (e.g. Cooper and Roberts, 2009; Cusack and Roberts, 2000; Dowling, 1973).

In the behavioral session, a random number of distractor sounds was played, in each trial, before and after the target stream (i.e. inducer sounds). It has been suggested that inducer sounds can trigger the build-up and therefore, may facilitate stream segregation (e.g. Roberts et al., 2008; Rogers and Bregman, 1993). To evaluate whether the presence of a random number of inducer sounds facilitated stream segregation, the d' scores from the behavioral full overlap condition were compared with those from the ERP recording session, where no inducer sounds were presented. The results from a mixed-effects linear model showed no significant effect of the inducer sounds [$F(1,11) = 2.701$, $p = 0.129$], indicating that these did not affect the d' scores.

Despite the similarity of the paradigms, the findings from the present study appear to be inconsistent with those reported by Cooper and Roberts (2009). In their study, the electrode separation between the streams did not affect the performance in the melody discrimination task. Moreover, most of their listeners performed near-chance level unless the distractor sounds were attenuated by at least 50% of the listener's dynamic range. In contrast, in the present study, all listeners were able to use electrode separation to segregate the streams when the target and the distractor stream were presented at the same loudness. Cooper and Roberts (2009) employed sequences with a fixed duration of 2.2 s whereas in the present study, the sequences had a duration which ranged between 4 and 6.65 s. Therefore, one might argue that in the study from Cooper and Roberts (2009), the listeners did not have enough time to build up a segregated percept, as suggested in chapter 2. If this was the case, in the present study, near-chance performance would be expected for the first deviant triplet, which happened between 1.3 and 2.6 s. Instead, the results from the behavioral experiment indicate that the effect of the electrode separation on the d' scores was largest for the first deviant triplet: The d' scores ranged from near-chance in the *full overlap* condition to near-ceiling in the *no overlap* condition.

In the study of Cooper and Roberts (2009) pitch direction judgments were required to identify the target melody. In contrast, in the present study, the listeners were not required to identify the melodic contour of the target stream, and could, instead, perform the task by detecting a change (i.e. deviant) in the target stream. Kong et al. (2004), suggested that as many as 32 independent frequency bands are needed to recognize familiar melodies. The spectral resolution of CIs is limited both by the number of electrodes (typically 22 or less) and the interaction of the electrical current across electrodes. Thus, current CIs may not provide enough spectral resolution to support melody recognition (e.g. Kong et al., 2004; Mehta and Oxenham, 2017). Therefore, the poor performance observed by Cooper and Roberts (2009) could reflect inherent limitations of CI listeners to recognize familiar melodies and may not be related to poor stream segregation.

4.4.2 The effect of selective attention on the event-related potentials

The N1 responses to the target and the distractor sounds were recorded while the listeners performed the behavioral task (i.e. active listening) and when the listeners watched a muted movie (i.e. passive listening). Even though the same physical stimuli were presented in both conditions, at the group level and when averaged across all deviant conditions, the N1 amplitudes were different in the active and in the passive listening conditions. Thus, selective auditory attention modulated the ERPs. Consistent with previous work in NH and HI listeners, the N1 responses to the target sounds were enhanced when the listeners performed the behavioral task vs. when they passively listened to the sounds (Choi et al., 2014; Dai and Shinn-Cunningham, 2016; Dai et al., 2018). However, attention did not significantly modulate the responses to the distractor sounds. This is consistent with previous studies suggesting that hearing impairment may affect the ability to suppress irrelevant sounds (Dai et al., 2018; Petersen et al., 2017).

The effect of selective attention on the N1 responses was largest in the *late* deviant condition, where the listeners had to sustain selective attention throughout the full duration of the sequence. In this condition, the N1 responses to the target sounds were enhanced during the second and third triplets, but not during the first one. Conversely, the N1 responses to the distractor sounds were enhanced for the first triplet, but not for the second and third ones (figure 4.6A). Thus, whereas the attentional modulation of the N1 responses to the target sounds became more robust over time, the attentional modulation of the N1 responses to the distractor sounds diminished over time. These results suggest that CI listeners become more effective at selectively listening to the target stream over time, in agreement with previous reports with NH listeners (e.g.

Choi et al., 2014; Dai et al., 2018; Shinn-Cunningham and Best, 2008). Moreover, only the N1 responses to the first two sounds of the target stream were enhanced in the active listening condition (figure 4.6B). Given the design of the stimuli, only the first two sounds of each triplet were necessary to detect the deviant. Thus, these results suggest that the listeners may have generally relied on the first two target sounds of each triplet to perform the task.

It has been suggested that the use of a cue to direct the listeners' attention towards a specific attribute of the sound (e.g. location or pitch) leads to anticipatory modulation of the cortical responses (Hill and Miller, 2010; Lee et al., 2012). As a result, the responses evoked by a sound will be larger when its attributes match those of the cue than when there is a mismatch. However other studies have suggested that the inherent salience of sudden onsets may override this anticipatory modulation, leading to similar ERP responses in the active and the passive listening conditions (e.g. Dai et al., 2018). In the present study, the target stream was temporally predictable (i.e. the lagging stream) and the listeners were familiar with both the target and the distractor stream. Thus, one would expect either a suppression effect on the N1 response to the first sound of the sequences (i.e. a distractor sound) or a lack of any attentional modulation. Instead, a larger N1 response to the first sound was observed for the active than for the passive listening condition (figure 4.4 and figure 4.5). These results suggest that CI listeners might not be able to ignore the leading stream and therefore attend to the first sound of the sequence (i.e. a distractor sound). As a consequence, they need to switch their attention from the distractor to the target stream during the trial. Consequently, the N1 attentional modulation of the distractor stream is significantly different from zero during the first triplet whereas the N1 attentional modulation of the target stream is significantly different than zero during the second and third triplets.

4.4.3 The relation between behavioral performance and the N1 attentional modulation

It has been suggested that when attention is focused on a particular stimulus feature, all coherent perceptual features may be bond together forming a stream (Shamma et al., 2011; Shamma et al., 2013). Other studies have suggested that selective attention may operate at the level of auditory objects, even if attention is initially focused on a particular feature (e.g. Bressler et al., 2014;

Shinn-Cunningham, 2008). The presence of a salient perceptual feature in the target stream may, therefore, facilitate the process of object formation. A clear object representation might, in turn, make the process of selective attention more effective. Correspondingly, several studies have found a significant correlation between the attentional modulation of cortical responses and behavioral performance on a selective auditory attention task (e.g. Choi et al., 2014; Dai and Shinn-Cunningham, 2016; Dai et al., 2018).

In the present study, no significant correlation was found between the individual performance in the deviant detection task and the amount of N1 attentional modulation. This was the case for all electrode separation conditions (figure 4.7). The d' scores in the *no overlap* condition (figure 4.7A and B) were, overall, at ceiling. This limited the individual variability of the d' scores, presumably contributing to the non-significant correlation between the behavioral performance and the N1 attentional modulation. In order to avoid ceiling effects in the behavioral performance while ensuring that enough correct trials are available to estimate the N1 response amplitude, Choi et al. (2014) correlated the behavioral performance in a challenging task with the N1 attentional modulation recorded under a less challenging condition. Choi et al. (2014) found a significant correlation between the behavioral performance and the amount of N1 attentional modulation. In contrast, in the present study, no significant correlation was found between the behavioral performance in the *apical* and *basal overlap* conditions, where no ceiling effects were present, and the amount of N1 attentional modulation (figure 4.7C and D).

The lack of a significant correlation between the d' scores and the N1 attentional modulation does not necessarily imply the independence of these two measures. Instead, it might reflect the limitations imposed by the conditions and paradigm chosen for the ERP recordings. The d' scores from the ERP recordings (*no overlap* condition) were at ceiling for most listeners, suggesting that the task was not demanding. Thus, it is possible that some listeners could have achieved high d' scores in the task without selectively attending to the target stream (e.g. they could have switched their attention between the streams). If this was the case, the individual differences in the N1 attentional modulation would not reflect differences in the ability to perceptually group the sounds. Moreover, the ERP recordings took place in the last of the three sessions. As a result, all listeners were familiar with the target stream at the time of the ERP recordings. During the passive listening condition, the listeners watched a muted movie and after the recording session, the listeners were asked informal questions about the movies. However, the individual level of engagement in the movie was not quantified. Thus, some listeners might have been engaged in the movie whereas other might have attended to the sound sequences. As a result, the N1 attentional modulation estimates from the present study might not be precise at an individual level.

4.5 Summary and conclusion

The present study combined a behavioral deviant detection task with ERP recordings to investigate the role of electrode separation in voluntary stream segregation for CI listeners. The results suggested that CI listeners can voluntarily segregate streams composed of multiple and different sounds when only electrode separation cues are provided. Moreover, a two-stream percept was found to build up over time. The results from the ERP recordings showed that auditory selective attention modulates the cortical responses in CI listeners. Specifically, selective attention enhanced the responses to the target sounds whereas responses to the distractor sounds remained unchanged. However, no correlation was found between the behavioral performance in the detection task and the attentional modulation of the ERPs.

Overall, the results from the present study suggest that CI listeners can use electrode separation to perceptually group sequentially presented sounds into auditory objects. Moreover, the effects of selective attention could be measured in CI listeners using EEG recordings (at the group level). This suggests that CI listeners, like NH listeners, are able to selectively attend to a target auditory object as long as its distinctive feature is sufficiently salient. The results from the present study also suggest that CI listeners might experience limitations in their ability to ignore a competing stream and to initially select the stream of interest, which might contribute to their poor performance in complex listening scenarios.

Acknowledgments

We would like to thank all volunteers who participated in this study. We thank Colette McKay for her help, advice and constructive criticism during the design of the experiments, Darren Mao for his help with the EEG setup and our colleagues from the Hearing Systems Group and the Translational Hearing Group for valuable comments and stimulating discussions. We thank Cochlear Ltd for providing us with some of the research equipment for this study.

Supplementary material

Cochlear implant artifact attenuation

The cochlear implant (CI) artifact was attenuated using a custom implementation of the procedure proposed by Viola et al. (2012). This approach makes use of independent component analysis (ICA) to decompose the electroencephalographic (EEG) signal into statistically maximally independent components. Instead of manually selecting the independent components (ICs), Viola et al. (2012) proposed three criteria to distinguish between ICs representing the neural activity, such as the N1 response, and the CI artifact: 1] The topography of ICs representing neural activity is well modeled with a dipole. Thus, the residual variance (RV) between the projection of the equivalent dipole model and the actual topography of the IC is generally low. Conversely, ICs representing the CI artifact exhibit less dipolar topographies and thus, a larger RV. 2] The largest activity of the N1 response is generally observed about 100 ms after the onset of the stimuli. Conversely, the largest activity of the CI artifact happens at the onset and offset of the stimulation. 3] ICs representing the artifact of a particular listener generally present similar topographies.

Based on these three criteria, Viola et al. (2012) proposed an algorithm to identify the ICs which represent the CI artifact. The algorithm consists of three steps. First, ICs with RV above a threshold are selected. Second, the first derivative of the selected ICs is calculated. The ratio between the root mean square (RMS) amplitude in the stimulus onset/offset time window and in the time window where the response of interest is expected (e.g. N1) is computed. The IC with the largest ratio is chosen as the topographical template for the CI artifact. The topography of the template is then correlated with the remaining ICs. In the third step, ICs either exceeding a ratio-threshold or a correlation-threshold are selected as CI components.

The algorithm proposed by Viola et al. (2012) was designed to attenuate the CI artifact from EEG responses to single sounds which could overlap with the time window of the neural response. In the present study each trial consisted on

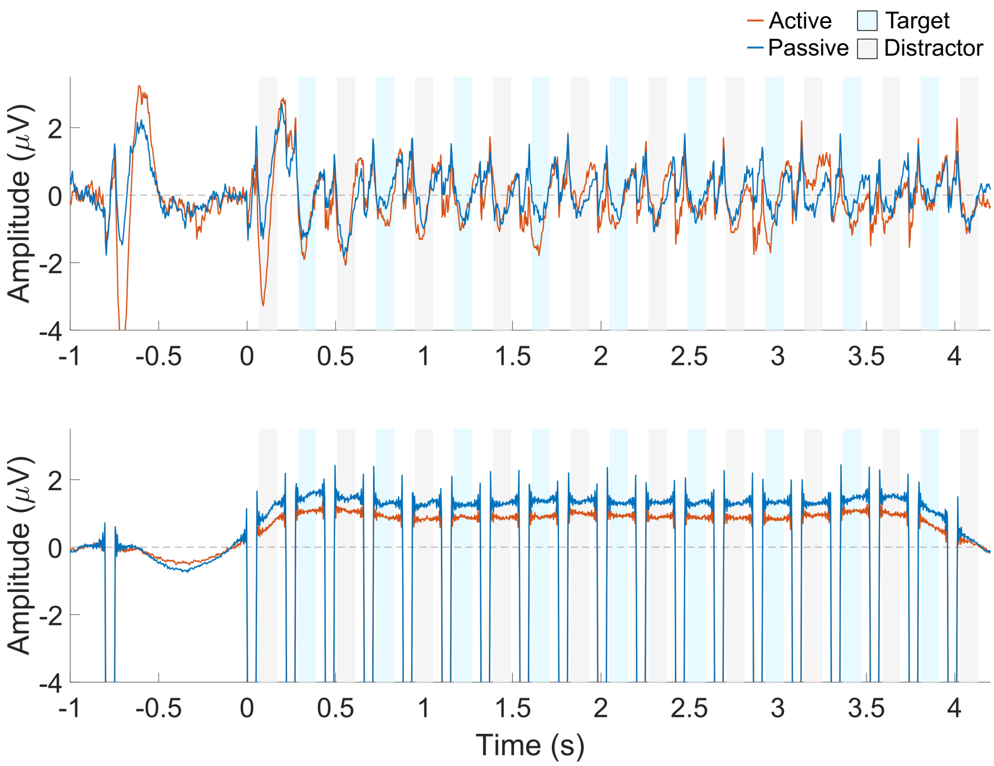


Figure 4.8: Grand average waveform for the clean data (top) and the artefactual data (bottom). The active listening condition is shown in red and the passive listening condition in blue. The blue and gray shaded areas indicate the N1 ERP component time window for the target and the distractor sounds, respectively. Each trace represents the average across nine front-central electrodes (Fz, AFz, FCz, F1, F2, FC1, FC2, AF3, AF4).

19 single electrode stimuli, each with a duration of 50 ms. For this reason, the artifact onset and offset time windows were replaced by a single time window from -10 to +60 ms (relative to the sound onset). As a result, a total of 19 RMS ratios were calculated in the second step. These values were averaged to obtain a single measure of the ratio between CI artifact and response for each IC. The RV-threshold was set to 10%, the ratio-threshold to 2.7 and the correlation –threshold to 0.85.

The algorithm was applied for each listener and listening condition (i.e. active and passive listening) independently. Thus, the artifact attenuation process could have introduced variations in the EEG waveforms that could be confounded with an attentional effect. To ensure that this was not the case, the ICs representing the artifact were back-projected to the sensor space (i.e. artefactual data). The clean data (i.e. data from which ICs representing the artefact had been removed) and the artefactual data were then processed in the same way.

The grand average waveform for the clean and the artefactual data is shown in figure 4.8. Red and blue solid lines represent the active and the passive listening conditions, respectively. Blue and gray shaded areas indicate the N1 response time window for the target and the distractor sounds, respectively. For the clean data (top panel), most of the activity is observed in the N1 response time windows. Sharp peaks can still be seen in the clean data, just before the N1 response time windows. This indicates that the CI artifact was not totally removed by the algorithm. For the artefactual data (bottom panel), little activity is observed in the N1 time windows. Instead, large square pulses are observed just before the N1 response time window, representing the CI artifact. These results imply that the artifact attenuation process did not introduce significant variations in the EEG waveform.

Detailed results from the post hoc analysis of the behavioral experiment

Table 4.2: Results from the pairwise comparison between the d' scores achieved for each deviant triplet and electrode separation condition. Independent comparisons were performed for each electrode separation condition. Reported p-values have been corrected for multiple comparisons using the Tukey method for a family of 12 estimates.

Electrode separation	Deviant triplet	Estimate	df	t-ratio	p-value
No overlap	1-2	-0.005	41.86	-0.021	1
	1-3	-0.614	49.55	-2.677	0.268
	2-3	-0.609	37.20	-2.419	0.419
Apical overlap	1-2	-0.272	41.86	-1.126	0.992
	1-3	-1.210	49.55	-5.276	<0.001
	2-3	-0.937	37.20	-3.725	0.027
Basal Overlap	1-2	-0.219	41.86	-0.907	0.999
	1-3	-1.457	49.55	-6.354	<0.001
	2-3	-1.237	37.20	-4.917	0.001
Full overlap	1-2	0.235	41.86	0.974	0.998
	1-3	0.071	49.55	0.309	1
	2-3	-0.165	37.20	-0.654	1

Table 4.3: Results from the pairwise comparison between the d' scores achieved for each deviant triplet and electrode separation condition. Independent comparisons were performed for each deviant triplet condition. No stands for no overlap, Apcl for apical overlap, Bsal for basal overlap and Full for full overlap. Reported p-values have been corrected for multiple comparisons using the Tukey method for a family of 12 estimates.

Deviant triplet	Electrode separation	Estimate	df	t-ratio	p-value
1	No-Apcl	1.120	24.16	3.908	0.026
	No-Bsl	1.843	21.95	3.908	< 0.001
	No-Full	2.612	27.52	6.075	<0.001
	Apcl-Bsl	0.723	26.33	2.641	0.308
	Apcl-Full	1.493	25.03	5.311	<0.001
	Bsl-Full	0.769	20.67	2.436	0.426
2	No-Apcl	0.852	24.16	2.975	0.177
	No-Bsl	1.628	21.95	5.369	0.001
	No-Full	2.853	27.52	10.643	<0.001
	Apcl-Bsl	0.776	26.33	2.834	0.222
	Apcl-Full	2.000	25.03	7.117	<0.001
	Bsl-Full	1.224	20.67	3.876	0.032
3	No-Apcl	0.524	24.16	1.827	0.789
	No-Bsl	1.000	21.95	3.296	0.100
	No-Full	3.296	27.52	12.299	<0.001
	Apcl-Bsl	0.476	26.33	1.739	0.835
	Apcl-Full	2.773	25.03	9.866	<0.001
	Bsl-Full	2.297	20.67	7.272	<0.001

5 General discussion

In this thesis, performance-based behavioral measures of voluntary stream segregation were carried out in cochlear implant (CI) listeners to investigate the role of place and temporal cues in the perceptual organization of sounds. Additionally, the attentional modulation of event-related potentials (ERPs), a neural correlate of selective attention, was recorded to investigate the time course of selective attention and whether it can be used as an objective measure of voluntary stream segregation in CI listeners. In the following, the main findings of the different chapters will be summarized and discussed.

5.1 The role of perceptual differences in stream segregation

In Chapters 2 and 3, the performance in a delay detection task was measured to investigate the role of electrode separation and pulse rate differences in voluntary stream segregation. The stimuli used in the experiments consisted of sequences of alternating A and B bursts of pulses with a fixed pulse rate, stimulating different electrodes (chapter 2) or stimulating a fixed electrode with different pulse rates (chapter 3). The target stream (B) was temporally regular, whereas the distractor stream (A) was temporally irregular. Thus, detecting a small delay applied to the last B sound of a sequence was easier when the A and B sounds were perceptually segregated. It was hypothesized that if the listeners were able to perceptually segregate sequential sounds on the basis of difference in the place or the rate of stimulation, the delay detection performance would improve as the perceptual difference between the A and the B sounds increased. In addition, if CI listeners experience a build-up effect, detection performance should improve as the duration of the sequence increased. The results showed that delay detection performance improved by increasing either the electrode separation or the pulse rate difference between the A and the B sounds. This indicates that larger differences in the place or the rate of stimulation facilitate sequential stream segregation for CI listeners. Moreover, the delay detection performance improved by increasing the duration of the sequence of A and B sounds from about 1.2 s to 4 s, suggesting that a two-stream percept builds up over time. The increase in performance observed for the longer sequences could also arise from a better sensitivity to temporal irregularities in the long sequences, i.e., the listeners would be exposed to a larger number of regular B-to-B time intervals and therefore, could benefit from integrating information over a longer period of time. However, the delay detection performance did not depend on the sequence duration when the B sounds were presented in quiet (i.e., when no stream segregation was needed to perform the task), indicating that the sequence duration per se did not influence the detection task.

In addition to investigating the role of pulse rate differences in stream segregation, chapter 3 evaluated the hypothesis that stream segregation may be related to the degree of the perceptual difference between the sounds (Moore and Gockel, 2002, 2012). Data from a verbal attribute magnitude experiment (Lamping et al., 2018) were used to convert the physical differences in the place and the rate of stimulation from chapters 2 and 3 to differences in the pitch height, a common perceptual dimension related to both physical cues. Perceptual pitch height differences accounted for the results obtained on the basis of both electrode and pulse rate differences, supporting the hypothesis that stream segregation is related to the degree of the perceptual difference between the sounds. Overall, a perceptual difference equivalent to about 20% of the pitch difference between electrodes 11 and 22 was needed to segregate the streams (i.e., the fission boundary), regardless of whether it was elicited by varying the place or the rate of stimulation.

The results presented in chapters 2 and 3 are, however, inconsistent with the findings from Cooper and Roberts (2009), which suggested that CI listeners may not be able to use electrode separation to perceptually segregate a melodic contour from random, distractor sounds. In the experiments from Cooper and Roberts (2009), all sequences had a fixed duration of 2.2 s and therefore, as suggested in chapter 2, the poor performance in their task could reflect the need of longer sequences to build up segregated percept. Nevertheless, factors such as differences in the complexity of the tasks or the presence or absence of rhythmic cues could also have caused the discrepancies between the results. Therefore, chapter 4 investigated whether the findings from chapter 2 hold when the listeners perform a more complex task, i.e., when each stream consists of multiple and different sounds and when no rhythmic cues are provided.

In chapter 4, stream segregation was investigated using a deviant detection task. The target stream consisted of a sequential pattern of bursts of pulses presented on electrodes 9, 11 and 13 (i.e. a triplet). In each trial, three triplets were presented consecutively in the presence of interleaved, random distractor sounds. A deviant triplet was randomly introduced by reversing the order of the electrode activation sequence in one of the three triplets. The listeners were asked to detect sequences that contained a deviant triplet and to report its location within the sequence. Unlike in chapters 2 and 3, both the target and the distractor streams were composed of multiple and different sounds. Thus, the listeners had to integrate different sounds to form a representation of each of the streams and to maintain these representations segregated over time, resembling the task used by Cooper and Roberts (2009). The sequences used in chapter 4 were longer than those from Cooper and Roberts (2009) (4 – 6.65 s vs. 2.2 s) and the build-up effect was assessed by comparing the detection performance for the three different locations of the deviant triplet. The electrode separation between the streams was varied in a similar way as in the study from Cooper and Roberts (2009). Thus, if the poor performance observed by Cooper and Roberts (2009) was due to the short duration of their sequences, detection performance should improve by increasing the electrode separation between the streams when the deviant was presented late in the sequence. Conversely, the electrode separation should have little or no effect when the deviant was presented early in the sequence, since the listeners would not have enough time to build up a segregated percept. The results did not support this hypothesis: detection performance improved with increasing electrode separation between the streams, reaching a ceiling when the target and the distractor streams were spatially separated in the electrode array. This was also the case when the deviant triplet was presented early in the sequence. A build-up effect was observed for the conditions where there was one electrode of overlap between the target and the distractor streams, i.e., when the performance was neither at ceiling nor near chance-level. This indicates that CI listeners can use electrode separation information to segregate sequences composed of multiple and different sounds, even when no rhythmic cues are provided. Thus, chapter 4 extended the findings from chapter 2 beyond simple sequences of alternating sounds. The results suggest that the poor performance observed by Cooper and Roberts (2009) may not be related to poor stream segregation. Instead, it could reflect inherent limitations of CI listeners to recognize melodies.

All 19 CI listeners who volunteered to participate in one or more of the studies from this thesis were able to perform the detection/discrimination task. This suggests that most CI listeners may be able to segregate sequentially presented sounds. Nevertheless, the behavioral results varied considerably across-listeners, consistent with previous studies investigating stream segregation in CI listeners (e.g. Chatterjee et al., 2006; Cooper and Roberts, 2009; Hong and Turner, 2006; Marozeau et al., 2013). The large across-listener variability could reflect the diversity of the group of listeners regarding their age, years of experience with the implant, years of deafness before implantation or etiology. However, the studies presented in this thesis were not designed to investigate the effect of these factors on stream segregation.

Overall, chapters 2, 3 and 2 presented evidence from objective, performance-based measures, indicating that CI listeners can segregate sequences of sounds on the basis of perceptual differences elicited through changes in either the place or the rate of stimulation and that a two-stream percept builds up over time. This is consistent with previous work suggesting that CI listeners may experience some aspects of stream segregation (Böckmann-Barthel et al., 2014; Chatterjee et al., 2006; Duran et al., 2012; Hong and Turner, 2006, 2009; Innes-Brown et al., 2011; Marozeau et al., 2013; Tejani et al., 2017). Moreover, perceptual differences were elicited via direct stimulation of the CI, bypassing the listener's speech processor and resulting in a better control of the signal delivered to the listeners.

5.2 Event-related potentials as a neural indicator of selective attention

The segregation of sounds into auditory streams is only part of the cocktail party problem. In order to selectively listen to a specific source in the presence of others, the listener also needs to sustain selective attention to the stream of interest, while ignoring the others (McDermott, 2009). Moreover, the processes of object formation and selective attention are closely related (e.g. Shamma et al., 2011; Shamma et al., 2013). Problems to select the stream of interest or to suppress a competing stream may therefore occur even when the sounds are perceptually grouped into streams (e.g. Shinn-Cunningham and Best, 2008). In the experiments described in chapters 2, 3 and 4, the listeners were asked to perform a behavioral task which required them to selectively attend to a subset of the sounds (i.e., the target stream) while ignoring the remaining sounds (i.e. the distractor stream). As a consequence, the results from these experiments were affected both by the stream segregation and the selective attention abilities of the listeners.

The performance in the behavioral tasks from chapters 2, 3 and 4 improved by increasing the perceptual difference between the streams. This indicates that the listeners were able to use perceptual differences to segregate the sounds and to selectively attend to the target stream. In these experiments the listeners provided a behavioral response at the end of the trial. Thus, they could take time to evaluate the sounds they heard, or to mentally replay the sequence, before making a decision. As a result, these behavioral measures provide little information about the process of selective attention.

In chapter 4, ERP recordings were used to specifically investigate the process of selective attention. It has been suggested that selective attention enhances the amplitude of the N1 responses elicited by the attended sounds, while it suppresses those elicited by the ignored sounds (e.g. Choi et al., 2013; Hillyard et al., 1973; Picton and Hillyard, 1974). In chapter 4, N1 responses were recorded while the listeners performed the behavioral task (active listening) and while the listeners watched a silent movie with captions (passive listening). Thus, it was hypothesized that if CI listeners can selectively attend to the target stream, the corresponding N1 amplitudes would be larger in the active than in the passive listening condition. Conversely, N1 responses evoked by the distractor stream would present a smaller amplitude in the active than in the passive listening condition. The amplitude of the N1 responses evoked by the target stream was, overall, larger in the active than in the passive listening condition. Conversely, selective attention did not affect the amplitude of the N1 responses evoked by the distractor sounds. These findings are consistent with recent studies with hearing-impaired listeners (Dai et al., 2018; Petersen et al., 2017), and suggest that whereas CI listeners may be able to selectively attend to the target stream, their abilities to suppress the distractor stream may be limited. In chapters 2 and 3, shorter delays were chosen for the control condition (i.e. without the distractor stream) to avoid ceiling effects. This implies that the presence of a distractor stream was detrimental for the detection task, even when the distractor sounds were perceptually different from the target sounds. In chapters 2 and 3, it was argued that the detrimental effect of the distractor stream could reflect a slow build-up process. However, this could also reflect the limitations from CI listeners to suppress a competing stream.

The amount of N1 attentional modulation was not correlated with the individual performance in the deviant detection task. Thus, even though selective attention modulated the amplitude of the N1 responses at the group level, this could not be used as an objective measure of stream segregation in the individual listeners. Previous studies with normal-hearing (NH) and hearing-impaired listeners (Choi et al., 2014; Dai and Shinn-Cunningham, 2016; Dai et al., 2018) suggested that the amount of attentional modulation and the behavioral performance in a task involving selective auditory attention may be closely related. The lack of a significant modulation between these measures in chapter 4 may not imply the independence of these two measures in CI listeners. Instead, it may reflect limitations arising from the specific conditions and paradigm chosen in chapter 4.

The effects of attention observed in the results from chapter 4 were larger in the trials where the deviant was presented either in the last triplet or when it was absent (i.e., the trials where the listeners had to sustain selective attention during the full trial). In these trials, selective attention initially enhanced the N1 responses evoked by the distractor stream (i.e. the leading stream) and not those evoked by the target stream (i.e. the lagging stream), despite the fact that the listeners were instructed to attend to the target stream. Conversely, during the second and third triplets, selective attention enhanced the N1 responses evoked by the target stream whereas no significant attentional modulation was observed for those responses evoked by the distractor stream. This suggests that CI listeners initially attended to the distractor stream and switched their attention to the target stream during the first triplet. Thus, they might be able to perceptually group the sounds into streams, but fail to select the object of interest at the beginning of the trial. As a result, they might initially attend to the leading stream, that may be perceptually more salient due to exogenous, bottom-up attention (e.g. Dai et al., 2018; Shinn-Cunningham, 2008).

It has been hypothesized that challenges with object selection might arise when the listeners do not know which stimulus feature distinguishes the target from the distractor stream, or when the perceptual differences between the streams are not sufficiently salient (e.g. Shinn-Cunningham, 2008; Shinn-Cunningham and Best, 2008). Given that the listeners were familiar with the target and the distractor streams, the challenges with object selection experienced by the listeners could reflect that the perceptual differences elicited by varying the stimulation electrode may not be sufficiently salient (see McDermott, 2004, for a review).

Overall, chapters 2, 3 and 4 present evidence from behavioral measures and from ERP recordings indicating that CI listeners are able to selectively attend to a target auditory object. Nevertheless, they might experience difficulties to suppress a competing stream and to select the object of interest.

5.3 The cochlear implant at a cocktail party

The findings presented throughout this thesis suggest that CI listeners are able to solve two of the problems at a "cocktail party": they can use perceptual differences to group sequential sounds into streams and they can sustain selective attention to the stream of interest. Specifically, this was the case when perceptual differences between the sounds were elicited by varying either the place or the rate of the electrical stimulation and when the listeners' intention biased the perception towards segregation (i.e. voluntary stream segregation). Moreover, previous studies suggested that CI listeners can voluntarily segregate sounds on the basis of loudness differences (Cooper and Roberts, 2009; Marozeau et al., 2013), and that CI listeners experience some aspects of obligatory stream segregation on the basis of differences in the place or rate of stimulation (e.g. Chatterjee et al., 2006; Duran et al., 2012; Tejani et al., 2017). However, no previous study estimated the temporal coherence boundary in CI listeners (i.e. the threshold for obligatory stream segregation), and no solid evidence for a build-up effect under integration-promoting listening instructions (i.e. obligatory stream segregation) has been reported (Cooper and Roberts, 2009). Thus, it is unclear whether CI listeners experience all aspects of obligatory stream segregation. The absence of robust obligatory processes would force CI listeners to predominantly rely on top-down, attention-driven processes for the perceptual organization of sounds. These require a priori knowledge about the features which distinguish the streams (e.g. Bregman, 1990, chapter 1), which could increase the cognitive load needed to perceptually group the sounds. The absence of robust obligatory mechanisms could partially account for the challenges that CI listeners experience in complex listening scenarios.

The focus of this thesis was to investigate sequential stream segregation. However, complex listening scenarios involve both simultaneous and sequential grouping: at a specific time instant, the listener needs to determine which frequency components belong to each stream (i.e. simultaneous grouping). Then, sound events belonging to the same stream need to be grouped over time (i.e. sequential grouping). Previous studies investigating simultaneous stream segregation in CI listeners suggested that differences in the stimulation pulse rate cannot be used to perceptually segregate concurrent sounds (Carlyon et al., 2007). Instead, they may be segregated on the basis of large onset asynchronies or large changes in level. However, the effect of these cues was small for most of the listeners and the authors concluded that the benefit from this cues may be reduced in complex listening scenarios (Carlyon et al., 2007; Cooper and Roberts, 2010). Thus, even though CI listeners may benefit from perceptual differences to segregate sequential sounds, their performance in complex listening scenarios may be compromised by the challenges they might experience to segregate concurrent sounds.

Current sound processing strategies could also limit the performance of CI listeners in complex listening scenarios. They result in a large number of electrodes being activated at the same time, effectively limiting the place information available to the listeners. The implications for stream segregation were shown in a study by Tejani et al. (2017). They assessed stream segregation using pure tone stimuli, presented through the listener's speech processor, and using direct electrical stimulation of the implant. They found significantly better stream segregation when the stimuli were presented via direct electrical stimulation than when they were presented acoustically, via the listener's sound processor.

Besides the limitations in the process of object formation, the analysis of the ERP recordings in chapter 4 suggested that even though CI listeners can selectively attend to the target stream, they might experience difficulties to suppress a competing stream and to initially select the object of interest. Complex listening scenarios, such as a conversation with multiple speakers, often require attention to switch rapidly from one object to another. Therefore, it has been suggested that problems in the object selection are likely to result in listeners missing part of the message (e.g. Shinn-Cunningham and Best, 2008). In that situation, the listeners may use knowledge about the spectrotemporal continuity of the signal as well as linguistic expectations to fill in the missing information (e.g. Bashford and Warren, 1987; Samuel, 2001), thereby increasing the cognitive load and the effort required to solve a complex auditory scene (e.g. Shinn-Cunningham and Best, 2008).

5.4 Perspectives

The scope of the studies presented here was to investigate the role of changes in the place and the rate of the electrical stimulation in voluntary stream segregation. Therefore, simple stimuli, where a single electrode was active at a given time instant, were considered. Further studies are needed to determine whether the conclusions presented here generalize to more complex stimuli (i.e. when multiple electrodes are stimulated simultaneously). In addition, the studies presented here used short bursts of pulses with relatively long offset-to-onset intervals. Further studies may investigate the role of the degree of temporal continuity between successive sounds by varying the offset-to-onset interval duration.

Previous studies showed that NH and CI listeners may benefit from the combination of perceptual cues across different sensory modalities (multimodal stimulation) and from context information for the perceptual segregation of sounds (e.g. Innes-Brown et al., 2011; Marozeau et al., 2010; Rahne et al., 2007; Slater and Marozeau, 2016). However, it is still unclear which auditory cues should be encoded via a different modality, or whether the findings from previous studies would generalize to complex auditory scenes. The studies described in this thesis were limited to the auditory modality, and further work is needed to understand the role of multimodal and context information for the perceptual organization of sounds in complex auditory scenes.

In NH listeners, auditory stream segregation has been assessed using both behavioral and electrophysiological measures. However, previous studies assessing stream segregation abilities in CI listeners focused exclusively on behavioral measures. Studies making use of electrophysiological measures to investigate stream segregation in NH listeners have contributed significantly to our understanding of the perceptual organization of sounds. Thus, future studies may consider electrophysiological measures to further investigate stream segregation in CI listeners. For example, change-detection responses, such as the MMN or the P300, could be used to better understand obligatory grouping processes in CI listeners.

Evidence from recent studies suggests that low-frequency cortical activity synchronizes to the attended auditory objects, allowing selective attention to be reliably decoded from electrophysiological recordings (e.g. O'Sullivan et al., 2014; Zion Golumbic et al., 2013). This has inspired researchers to work towards the development of brain computer interfaces to steer a hearing aid using brain signals (e.g. Cheveigné et al., 2018; Fuglsang et al., 2017). The results presented in chapter 4 showed that the effects of selective attention can be measured in CI listeners using electrophysiological recordings of cortical activity. Moreover, chapters 2, 3 and 4 suggest that CI listeners may benefit from larger perceptual differences between the target and the distractor sounds. Thus, it may be possible to decode attention from brain signals and to use this information to enhance the perceptual difference between the target stream and the background sounds. However, further studies are needed to determine whether CIs, like hearing aids, can be steered by brain signals.

Finally, computational models are a powerful tool to deepen our understanding about the perceptual organization of sounds. Despite the large number of computational models of auditory scene analysis (for a review, see Szabó et al., 2016; Wang and Brown, 2006), little attention has been given to them as a tool to better understand the perceptual organization of sounds by CI listeners. The behavioral data published with this thesis could be used to validate a future computational model of auditory scene analysis for electric hearing.

5.5 Conclusions

The results presented throughout the chapters of this thesis indicate that:

- CI listeners can use perceptual differences elicited by varying the place or the rate of the electrical stimulation to voluntarily segregate sequential sounds. Stream segregation is related to the degree of perceptual difference between the sounds and a segregated percept generally builds up over time.

- These findings are not limited to sequences of two alternating sounds and they apply to more complex stimuli where each stream consists of multiple and different sounds.

- CI listeners can selectively attend to a target stream. Selective auditory attention enhanced the amplitude of the ERPs elicited by the attended sounds (at the group level). However, CI listeners were found to experience difficulties to suppress a competing stream and to initially select the stream of interest.

Overall, these findings represent a valuable basis for future studies investigating the perceptual organization of sounds by CI listeners, for the development of future CIs, which may be steered using brain signals, and for the validation of future computational models of auditory scene analysis for electric hearing.

Bibliography

Alain, C., S. R. Arnott, and B. J. Dyson (2015). "Varieties of Auditory Attention". In: *Oxford Handb. Cogn. Neurosci.* Vol. 1. January, pp. 1–24.

Anstis, S. M. and S. Saida (1985). "Adaptation to auditory streaming of frequency-modulated tones." In: *J. Exp. Psychol. Hum. Percept. Perform.* 11.5, pp. 257–271.

Bashford, J. A. and R. M. Warren (1987). "Multiple phonemic restorations follow the rules for auditory induction". In: *Percept. Psychophys.* 42.2, pp. 114–121.

Bates, D., M. Mächler, B. Bolker, and S. Walker (2014). "Fitting linear mixed-effects models using lme4". In: *J. Stat. Softw.* 67.1, p. 51.

Baumann, U. and A. Nobbe (2004). "Pulse rate discrimination with deeply inserted electrode arrays". In: *Hear. Res.* 196.1-2, pp. 49–57.

Bendixen, A. (2014). "Predictability effects in auditory scene analysis: a review". In: *Front. Neurosci.* 8.March, pp. 1–16.

Benjamini, Y. and Y. Hochberg (1995). "Controlling the false discovery rate : a practical and powerful approach to multiple testing". In: *J R Stat. Soc B* 57.1, pp. 289–300.

Böckmann-Barthel, M., S. Deike, A. Brechmann, M. Ziese, and J. L. Verhey (2014). "Time course of auditory streaming: do CI users differ from normal-hearing listeners?" In: *Front. Psychol.* 5.July, p. 775.

Bregman, A. S. (1978). "Auditory streaming is cumulative". In: *J. Exp. Psychol. Hum. Percept. Perform.* 4.3, pp. 380–387.

Bregman, A. S. (1990). *Auditory scene analysis : The perceptual organization of sound.* The MIT Press.

Bregman, A. S. and J. Campbell (1971). "Primary auditory stream segregation and perception of order in rapid sequences of tones." In: *J. Exp. Psychol.* 89.2, pp. 244–249.

Bressler, S., S. Masud, H. Bharadwaj, and B. Shinn-Cunningham (2014). "Bottom-up influences of voice continuity in focusing selective auditory attention". In: *Psychol. Res.* 78.3, pp. 349–360.

Carlyon, R. P. (2004). “How the brain separates sounds”. In: *Trends Cogn. Sci.* 8.10, pp. 465–471.

Carlyon, R. P., C. J. Long, J. M. Deeks, and C. M. McKay (2007). “Concurrent sound segregation in electric and acoustic hearing”. In: *JARO - J. Assoc. Res. Otolaryngol.* 8.1, pp. 119–133.

Chatterjee, M, Q. J. Fu, and R. V. Shannon (1998). “Within-channel gap detection using dissimilar markers in cochlear implant listeners.” In: *J. Acoust. Soc. Am.* 103.5 Pt 1, pp. 2515–2519.

Chatterjee, M., A. Sarampalis, and S. I. Oba (2006). “Auditory stream segregation with cochlear implants: A preliminary report”. In: *Hear. Res.* 222.1-2, pp. 100–107.

Cherry, E. C. (1953). “Some Further Experiments upon the Recognition of Speech, with One and with Two Ears”. In: *J. Acoust. Soc. Am.* 25.5, pp. 975–979.

Cheveigné, A. de, D. E. Wong, G. M. Di Liberto, J. Hjortkjær, M. Slaney, and E. Lalor (2018). “Decoding the auditory brain with canonical component analysis”. In: *Neuroimage* 172.September 2017, pp. 206–216.

Choi, I., S. Rajaram, L. a. Varghese, and B. G. Shinn-Cunningham (2013). “Quantifying attentional modulation of auditory-evoked cortical responses from single-trial electroencephalography.” In: *Front. Hum. Neurosci.* 7.April, p. 115.

Choi, I., L. Wang, H. Bharadwaj, and B. Shinn-Cunningham (2014). “Individual differences in attentional modulation of cortical responses correlate with selective attention performance”. In: *Hear. Res.* 314, pp. 10–19.

Christiansen, S. K., M. L. Jepsen, and T. Dau (2014). “Effects of tonotopicity, adaptation, modulation tuning, and temporal coherence in "primitive" auditory stream segregation.” In: *J. Acoust. Soc. Am.* 135.1, pp. 323–33.

Collins, L. M., T. a. Zwolan, and G. H. Wakefield (1997). “Comparison of electrode discrimination, pitch ranking, and pitch scaling data in postlingually deafened adult cochlear implant subjects.” In: *J. Acoust. Soc. Am.* 101.1, pp. 440–55.

Cooper, H. R. and B. Roberts (2007). “Auditory stream segregation of tone sequences in cochlear implant listeners”. In: *Hear. Res.* 225.1-2, pp. 11–24.

Cooper, H. R. and B. Roberts (2009). “Auditory stream segregation in cochlear implant listeners: measures based on temporal discrimination and interleaved melody recognition.” In: *J. Acoust. Soc. Am.* 126.4, pp. 1975–1987.

Cooper, H. R. and B. Roberts (2010). “Simultaneous grouping in cochlear implant listeners: Can abrupt changes in level be used to segregate components from a complex tone?” In: *JARO - J. Assoc. Res. Otolaryngol.* 11.1, pp. 89–100.

Cusack, R. and B Roberts (2000). “Effects of differences in timbre on sequential grouping.” In: *Percept. Psychophys.* 62.5, pp. 1112–1120.

Dai, L. and B. G. Shinn-Cunningham (2016). “Contributions of Sensory Coding and Attentional Control to Individual Differences in Performance in Spatial Auditory Selective Attention Tasks”. In: *Front. Hum. Neurosci.* 10.October, pp. 1–19.

Dai, L., V. Best, and B. G. Shinn-Cunningham (2018). “Sensorineural hearing loss degrades behavioral and physiological measures of human spatial selective auditory attention”. In: *Proc. Natl. Acad. Sci.* P. 201721226.

David, M., N. Grimault, and M. Lavandier (2015). “Sequential streaming, binaural cues and lateralization”. In: *J. Acoust. Soc. Am.* 138.6, pp. 3500–3512.

Dowling, W. J. (1973). “The perception of interleaved melodies”. In: *Cogn. Psychol.* 5.3, pp. 322–337.

Duran, S. I., L. M. Collins, and C. S. Throckmorton (2012). “Stream segregation on a single electrode as a function of pulse rate in cochlear implant listeners.” In: *J. Acoust. Soc. Am.* 132.6, pp. 3849–3855.

Eddington, D. K., W. H. Dobelle, D. E. Brackmann, M. G. Mladejovsky, and J Parkin (1978). “Place and periodicity pitch by stimulation of multiple scala tympani electrodes in deaf volunteers.” In: *Trans. Am. Soc. Artif. Intern. Organs* 24, pp. 1–5.

Fuglsang, S. A., T. Dau, and J. Hjortkjær (2017). “Noise-robust cortical tracking of attended speech in real-world acoustic scenes”. In: *Neuroimage* 156.December 2016, pp. 435–444.

Gfeller, K. et al. (2005). “Recognition of "real-world" musical excerpts by cochlear implant recipients and normal-hearing adults”. In: *Ear Hear.* 26.3, pp. 237–250.

Grimault, N., S. P. Bacon, and C. Micheyl (2002). “Auditory stream segregation on the basis of amplitude-modulation rate.” In: *J. Acoust. Soc. Am.* 111.3, pp. 1340–1348.

Gutschalk, A. (2005). “Neuromagnetic Correlates of Streaming in Human Auditory Cortex”. In: *J. Neurosci.* 25.22, pp. 5382–5388.

Gutschalk, A. and A. R. Dykstra (2014). “Functional imaging of auditory scene analysis.” In: *Hear. Res.* 307, pp. 98–110.

Hanekom, J. J. and R. V. Shannon (1998). "Gap detection as a measure of electrode interaction in cochlear implants." In: *J. Acoust. Soc. Am.* 104.4, pp. 2372–2384.

Hautus, M. J. (1995). "Corrections for extreme proportions and their biasing effects on estimated values of d'". In: *Behav. Res. Methods, Instruments, Comput.* 27.1, pp. 46–51.

Hill, K. T. and L. M. Miller (2010). "Auditory attentional control and selection during cocktail party listening". In: *Cereb. Cortex* 20.3, pp. 583–590.

Hill, K. T., C. W. Bishop, and L. M. Miller (2012). "Auditory grouping mechanisms reflect a sound's relative position in a sequence". In: *Front. Hum. Neurosci.* 6.June, pp. 1–7.

Hillyard, S. a., E. K. Vogel, and S. J. Luck (1998). "Sensory gain control (amplification) as a mechanism of selective attention: electrophysiological and neuroimaging evidence." In: *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 353.1373, pp. 1257–1270.

Hillyard, S, R Hink, V Schewent, and T Picton (1973). "Electrical signs of selective attention in the human brain". In: *Science (80-.).* 182.4108, pp. 177–180.

Hoesel, R. J. van and G. M. Clark (1997). "Psychophysical studies with two binaural cochlear implant subjects." In: *J. Acoust. Soc. Am.* 102.1, pp. 495–507.

Hong, R. S. and C. W. Turner (2006). "Pure-tone auditory stream segregation and speech perception in noise in cochlear implant recipients." In: *J. Acoust. Soc. Am.* 120.1, pp. 360–374.

Hong, R. S. and C. W. Turner (2009). "Sequential stream segregation using temporal periodicity cues in cochlear implant recipients." In: *J. Acoust. Soc. Am.* 126.1, pp. 291–299.

Hupé, J.-M. and D. Pressnitzer (2012). "The initial phase of auditory and visual scene analysis." In: *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 367.1591, pp. 942–53.

Innes-Brown, H, J Marozeau, and P Blamey (2011). "The effect of visual cues on difficulty ratings for segregation of musical streams in listeners with impaired hearing". eng. In: *PLoS One* 6.12, e29327.

Iverson, P (1995). "Auditory stream segregation by musical timbre: effects of static and dynamic acoustic attributes." In: *J. Exp. Psychol. Hum. Percept. Perform.* 21.4, pp. 751–763.

Koffka, K. (1935). *Principles of Gestalt Psychology.* Harcourt, Brace.

Kong, Y.-Y., R. Cruz, J. A. Jones, and F.-G. Zeng (2004). "Music perception with temporal cues in acoustic and electric hearing". In: *Ear Hear.* 25.2, pp. 173–185.

Kuznetsova, A., R. H. Christensen, C. Bavay, and P. B. Brockhoff (2015). "Automated mixed ANOVA modeling of sensory and consumer data". In: *Food Qual. Prefer.* 40, pp. 31–38.

Kuznetsova, A., P. B. Brockhoff, and R. H. B. Christensen (2017). "lmerTest Package: Tests in Linear Mixed Effects Models". In: *J. Stat. Softw.* 82.13, pp. 1 –26.

Lamping, W., S. Santurette, and J. Marozeau (2018). "Verbal attribute magnitude estimates of pulse trains across electrode places and stimulation rates in cochlear implant listeners". In: *Proc. Int. Symp. Audit. Audiol. Res. Vol 6 Adapt. Process. Hear.* Vol. 6.

Landsberger, D. M., K. Vermeire, A. Claes, V. Van Rompaey, and P. Van de Heyning (2016). "Qualities of single electrode stimulation as a function of rate and place of stimulation with a cochlear implant". In: *Ear Hear.* 37.3, e149–e159.

Lee, A. K. et al. (2012). "Auditory selective attention reveals preparatory activity in different cortical regions for selection based on source location and source pitch". In: *Front. Neurosci.* 6.JAN, pp. 1–9.

Lenth, R. V. (2016). "Least-squares means: The {R} package {lsmeans}". In: *J. Stat. Softw.* 69.1, pp. 1–33.

Long, C. J. et al. (2005). "Optimizing the clinical fit of auditory brain stem implants". In: *Ear Hear.* 26.3, pp. 251–262.

Luo, X., M. Padilla, and D. M. Landsberger (2012). "Pitch contour identification with combined place and temporal cues using cochlear implants". In: *J. Acoust. Soc. Am.* 131.2, pp. 1325–1336.

Macherey, O. and R. P. Carlyon (2010). "Temporal pitch percepts elicited by dual-channel stimulation of a cochlear implant." In: *J. Acoust. Soc. Am.* 127.1, pp. 339–349.

Mackersie, C. L., T. L. Prida, and D. Stiles (2001). "The role of sequential stream segregation and frequency selectivity in the perception of simultaneous sentences by listeners with sensorineural hearing loss". In: *J. Speech Lang. Hear. Res.* 44.February, pp. 19–28.

Marimuthu, V., B. A. Swanson, and R. Mannell (2016). "Cochlear Implant Rate Pitch and Melody Perception as a Function of Place and Number of Electrodes". In: *Trends Hear.* 20, pp. 1–20.

Marozeau, J., H. Innes-Brown, D. B. Grayden, A. N. Burkitt, and P. J. Blamey (2010). "The effect of visual cues on auditory stream segregation in musicians and non-musicians". In: *PLoS One* 5.6, pp. 1–10.

Marozeau, J., H. Innes-Brown, and P. J. Blamey (2013). "The acoustic and perceptual cues affecting melody segregation for listeners with a cochlear implant". In: *Front. Psychol.* 4.790, pp. 1–11.

McDermott, H. J. (2004). "Music perception with cochlear implants: a review". In: *Trends Amplif.* 8.2, pp. 49–82.

McDermott, J. H. (2009). "The cocktail party problem." In: *Curr. Biol.* 19.22, R1024–R1027.

McKay, C. M., H. J. McDermott, and R. P. Carlyon (2000). "Place and temporal cues in pitch perception: are they truly independent?" In: *Acoust. Res. Lett. Online* 1.1, p. 25.

Mehta, A. H. and A. J. Oxenham (2017). "Vocoder Simulations Explain Complex Pitch Perception Limitations Experienced by Cochlear Implant Users". In: *JARO - J. Assoc. Res. Otolaryngol.* 18.6, pp. 789–802.

Micheyl, C. and A. J. Oxenham (2010a). "Objective and subjective psychophysical measures of auditory stream integration and segregation". In: *JARO - J. Assoc. Res. Otolaryngol.* 11.4, pp. 709–724.

Micheyl, C. and A. J. Oxenham (2010b). "Pitch, harmonicity and concurrent sound segregation: Psychoacoustical and neurophysiological findings". In: *Hear. Res.* 266.1-2, pp. 36–51.

Micheyl, C., R. P. Carlyon, R. Cusack, and B. C. J. Moore (2005). "Performance measures of auditory organization". In: *Audit. Signal Process. Physiol. Psychoacoustics, Model.* Pp. 203–211.

Moore, B. C. and B. R. Glasberg (1988). "Gap detection with sinusoids and noise in normal, impaired, and electrically stimulated ears." In: *J. Acoust. Soc. Am.* 83.3, pp. 1093–1101.

Moore, B. C. J. and H. E. Gockel (2002). "Factors influencing sequential stream segregation". In: *Acta Acust. united with Acust.* 88.3, pp. 320–333.

Moore, B. C. J. and H. E. Gockel (2012). "Properties of auditory stream formation". In: *Philos. Trans. R. Soc. B Biol. Sci.* 367, pp. 919–931.

Näätänen, R., P. Paavilainen, T. Rinne, and K. Alho (2007). "The mismatch negativity (MMN) in basic research of central auditory processing: A review". In: *Clin. Neurophysiol.* 118.12, pp. 2544–2590.

Nelson, D. A., D. J. Van Tasell, A. C. Schroder, S Soli, and S Levine (1995). “Electrode ranking of place pitch and speech recognition in electrical hearing.” In: *J. Acoust. Soc. Am.* 98.4, pp. 1987–99.

Nelson, P. B., S.-H. Jin, A. E. Carney, and D. A. Nelson (2003). “Understanding speech in modulated interference: Cochlear implant users and normal-hearing listeners”. In: *J. Acoust. Soc. Am.* 113.2, pp. 961–968.

Nie, Y. and P. Nelson (2015). “Auditory stream segregation using amplitude modulated bandpass noise.” In: *J. Acoust. Soc. Am.* 127.3, p. 1809.

Nie, Y., Y. Zhang, and P. B. Nelson (2014). “Auditory stream segregation using bandpass noises: evidence from event-related potentials”. In: *Front. Neurosci.* 8.September, pp. 1–12.

O’Sullivan, J. a. et al. (2014). “Attentional Selection in a Cocktail Party Environment Can Be Decoded from Single-Trial EEG.” In: *Cereb. Cortex*, pp. 1–10.

Oostenveld, R., P. Fries, E. Maris, and J. M. Schoffelen (2011). “FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data”. In: *Comput. Intell. Neurosci.* 2011.

Oxenham, A. J. (2008). “Pitch perception and auditory stream segregation: implications for hearing loss and cochlear implants”. In: *Trends Amplif.* 12.4, pp. 316–331.

Oxenham, A. J. (2018). “How We Hear: The Perception and Neural Coding of Sound”. In: *Annu. Rev. Psychol.* 69.1, pp. 27–50.

Paredes-Gallardo, A., S. M. K. Madsen, T. Dau, and J. Marozeau (2018a). “The role of place cues in voluntary stream segregation for cochlear implant users”. In: *Trends Hear.* 22, pp. 1–13.

Paredes-Gallardo, A., S. M. Madsen, T. Dau, and J. Marozeau (2018b). “The role of temporal cues on voluntary stream segregation in cochlear implant users”. In: *Trends Hear.* 22, pp. 1–13.

Petersen, E. B., M. Wöstmann, J. Obleser, and T. Lunner (2017). “Neural tracking of attended versus ignored speech is differentially affected by hearing loss”. In: *J. Neurophysiol.* 117.1, pp. 18–27.

Picton, T. W. and S. A. Hillyard (1974). “Human auditory evoked potentials. II: effects of attention”. In: *Electroencephalogr. Clin. Neurophysiol.* 36, pp. 191–199.

Pressnitzer, D. and J.-M. Hupé (2006). “Temporal Dynamics of Auditory and Visual Bistability Reveal Common Principles of Perceptual Organization”. In: *Curr. Biol.* 16.13, pp. 1351–1357.

Qi, S. and R. E. Mitchell (2012). “Large-scale academic achievement testing of deaf and hard-of-hearing students: Past, present, and future”. In: *J. Deaf Stud. Deaf Educ.* 17.1, pp. 1–18.

Rader, T., J. Döge, Y. Adel, T. Weissgerber, and U. Baumann (2016). “Place dependent stimulation rates improve pitch perception in cochlear implantees with single-sided deafness”. In: *Hear. Res.* 354, p. 109.

Rahne, T, M Bockmann, H von Specht, and E. S. Sussman (2007). “Visual cues can modulate integration and segregation of objects in auditory scene analysis”. In: *Brain Res* 1144, pp. 127–135.

Rajendran, V. G., N. S. Harper, B. D. Willmore, W. M. Hartmann, and J. W. H. Schnupp (2013). “Temporal predictability as a grouping cue in the perception of auditory streams”. In: *J. Acoust. Soc. Am.* 134.1, EL98.

Roberts, B., B. R. Glasberg, and B. C. J. Moore (2002). “Primitive stream segregation of tone sequences without differences in fundamental frequency or passband.” In: *J. Acoust. Soc. Am.* 112.5 Pt 1, pp. 2074–2085.

Roberts, B., B. R. Glasberg, and B. C. J. Moore (2008). “Effects of the build-up and resetting of auditory stream segregation on temporal discrimination.” In: *J. Exp. Psychol. Hum. Percept. Perform.* 34.4, pp. 992–1006.

Rogers, W. L. and A. S. Bregman (1993). “An experimental evaluation of three theories of auditory stream segregation.” In: *Percept. Psychophys.* 53.2, pp. 179–189.

Rose, M. M. and B. C. J. Moore (2005). “The relationship between stream segregation and frequency discrimination in normally hearing and hearing-impaired subjects”. In: *Hear. Res.* 204.1-2, pp. 16–28.

Sach, A. J. and P. J. Bailey (2004). “Some characteristics of auditory spatial attention revealed using rhythmic masking release.” In: *Percept. & psychophys.* 66.8, pp. 1379–1387.

Samuel, A. G. (2001). “Knowing a word affects the fundamental perception of the sounds within it”. In: *Psychol. Sci.* 12.4, pp. 348–351.

Shamma, S. a., M. Elhilali, and C. Micheyl (2011). “Temporal coherence and attention in auditory scene analysis.” In: *Trends Neurosci.* 34.3, pp. 114–23.

Shamma, S. et al. (2013). “Temporal Coherence and the Streaming of Complex Sounds”. English. In: *BASIC Asp. Hear. Physiol. Percept.* 787. Ed. by B. C. J. Moore Patterson, RD, Winter, IM, Carlyon, RP, Gockel, HE, pp. 535–543.

Shannon, R. V. (1989). “Detection of gaps in sinusoids and pulse trains by patients with cochlear implants.” In: *J. Acoust. Soc. Am.* 85.6, pp. 2587–2592.

Shannon, R. V. (1992). “Temporal modulation transfer functions in patients with cochlear implants”. In: *J. Acoust. Soc. Am.* 91.4 Pt 1, pp. 2156–2164.

Shannon, R. (1983). “Multichannel electrical stimulation of the auditory nerve in man. I. Basic psychophysics”. In: *Hear. Res.* 11.2, pp. 157–189.

Shannon, R., F.-G. Zeng, V. Kamath, J. Wygonsky, and M. Ekelid (1995). “Speech recognition with primarily temporal cues”. In: *Science (80-.).* 270.5234, pp. 303–304.

Shinn-Cunningham, B. G. (2008). “Object-based auditory and visual attention”. In: *Trends Cogn. Sci.* 12.5, pp. 182–186.

Shinn-Cunningham, B. G. and V. Best (2008). “Selective Attention in Normal and Impaired Hearing”. In: *Trends Amplif.* Pp. 1–17.

Singh, P. G. and A. S. Bregman (1997). “The influence of different timbre attributes on the perceptual segregation of complex-tone sequences.” In: *J. Acoust. Soc. Am.* 102.4, pp. 1943–1952.

Slater, K. D. and J. Marozeau (2016). “The effect of tactile cues on auditory stream segregation ability of musicians and nonmusicians.” In: *Psychomusicology Music. Mind, Brain* 26.2, pp. 162–166.

Snyder, J., C. Alain, and T. Picton (2006). “Effects of attention on neuroelectric correlates of auditory stream segregation”. In: *J. Cogn. Neurosci.* Pp. 1–13.

Snyder, J. S. and C. Alain (2007). “Toward a neurophysiological theory of auditory stream segregation.” In: *Psychol. Bull.* 133.5, pp. 780–799.

Stainsby, T. H., C. Fu, H. J. Flanagan, S. K. Waldman, and B. C. J. Moore (2011). “Sequential streaming due to manipulation of interaural time”. In: *J. Acoust. Soc. Am.* 130.2, pp. 904–914.

Sussman, E, W Ritter, and H. G. Vaughan (1999). “An investigation of the auditory streaming effect using event-related brain potentials.” In: *Psychophysiology* 36.1, pp. 22–34.

Sutton, S., M. Braren, J. Zubin, and E. R. John (1965). “Evoked-Potential Correlates of Stimulus Uncertainty”. In: *Science (80-.).* 150.3700, pp. 1187–1188.

Szabó, B. T., S. L. Denham, and I. Winkler (2016). “Computational models of auditory scene analysis: A review”. In: *Front. Neurosci.* 10.NOV, pp. 1–16.

Tejani, V. D., K. C. Schvartz-Leyzac, and M. Chatterjee (2017). "Sequential stream segregation in normally-hearing and cochlear-implant listeners". In: *J. Acoust. Soc. Am.* 141.1, pp. 50–64.

Thompson, S. K., R. P. Carlyon, and R. Cusack (2011). "An objective measurement of the build-up of auditory streaming and of its modulation by attention." In: *J. Exp. Psychol. Hum. Percept. Perform.* 37.4, pp. 1253–1262.

Tomblin, J. B., M. Harrison, S. E. Ambrose, E. A. Walker, J. J. Oleson, and M. P. Moeller (2015). "Language outcomes in young children with mild to severe hearing loss". In: *Ear Hear.* Pp. 76–91.

Tong, Y. C. and G. M. Clark (1985). "Absolute identification of electric pulse rates and electrode positions by cochlear implant patients." In: *J. Acoust. Soc. Am.* 77.5, pp. 1881–8.

Tong, Y. C., G. M. Clark, P. J. Blamey, P. A. Busby, and R. C. Dowell (1980). "Psychophysical studies on two multiple channel cochlear implant patients". In: *J. Acoust. Soc. Am.* 71.November 1980, pp. 153–160.

Tong, Y. C., P. J. Blamey, R. C. Dowell, and G. M. Clark (1983). "Psychophysical studies evaluating the feasibility of a speech processing strategy for a multiple-channel cochlear implant." In: *J. Acoust. Soc. Am.* 74.1, pp. 73–80.

Townshend, B., N. Cotter, D. V. Compernolle, and R. L. White (1987). "Pitch perception by cochlear implant subjects". In: *J. Acoust. Soc. Am.* 82.1, pp. 106–115.

Van Noorden, L. (1977). "Minimum differences of level and frequency for perceptual fission of tone sequences ABAB*". In: *J. Acoust. Soc. Am.* 61.4, pp. 1041–1045.

Van Noorden, L. P. A. S. (1975). "Temporal coherence in the perception of tone sequences". PhD thesis.

Verde, M. F., N. A. Macmillan, and C. M. Rotello (2006). "Measures of sensitivity based on a single hit rate and false alarm rate: The accuracy, precision, and robustness of,A z, andA'". In: *Percept. Psychophys.* 68.4, pp. 643–654.

Viola, F. C. et al. (2012). "Semi-automatic attenuation of cochlear implant artifacts for the evaluation of late auditory evoked potentials". In: *Hear. Res.* 284.1-2, pp. 6–15.

Vliegen, J and a. J. Oxenham (1999). "Sequential stream segregation in the absence of spectral cues." In: *J. Acoust. Soc. Am.* 105.October 1997, pp. 339–346.

Vliegen, J, B. C. J. Moore, and A. J. Oxenham (1999). “The role of spectral and periodicity cues in auditory stream segregation, measured using a temporal discrimination task.” In: *J. Acoust. Soc. Am.* 106.2, pp. 938–945.

Wang, D. and G. J. Brown (2006). *Computational Auditory Scene Analysis: Principles, Algorithms, and Applications.* eng. Wiley-IEEE Press.

Wieringen, a van and J Wouters (1999). “Gap detection in single- and multiple-channel stimuli by LAURA cochlear implantees.” In: *J. Acoust. Soc. Am.* 106.4 Pt 1, pp. 1925–1939.

Wilson, B. S., D. L. Tucci, M. H. Merson, and G. M. O’Donoghue (2017). “Global hearing health care: new findings and perspectives”. In: *Lancet* 390.10111, pp. 2503–2515.

Zeng, F. G. (2002). “Temporal pitch in electric hearing”. In: *Hear. Res.* 174.1-2, pp. 101–106.

Zeng, F. G., S. Rebscher, W. Harrison, X. Sun, and H. Feng (2008). “Cochlear implants: system design, integration, and evaluation.” In: *IEEE Rev. Biomed. Eng.* 1.dc, pp. 115–142.

Zion Golumbic, E. M. et al. (2013). “Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party".” In: *Neuron* 77.5, pp. 980–91.

Contributions to Hearing Research

Vol. 1: *Gilles Pigasse*, Deriving cochlear delays in humans using otoacoustic emissions and auditory evoked potentials, 2008.

Vol. 2: *Olaf Strelcyk*, Peripheral auditory processing and speech reception in impaired hearing, 2009.

Vol. 3: *Eric R. Thompson*, Characterizing binaural processing of amplitude-modulated sounds, 2009.

Vol. 4: *Tobias Piechowiak*, Spectro-temporal analysis of complex sounds in the human auditory system, 2009.

Vol. 5: *Jens Bo Nielsen*, Assessment of speech intelligibility in background noise and reverberation, 2009.

Vol. 6: *Helen Connor*, Hearing aid amplification at soft input levels, 2010.

Vol. 7: *Morten Løve Jepsen*, Modeling auditory processing and speech perception in hearing-impaired listeners, 2010.

Vol. 8: *Sarah Verhulst*, Characterizing and modeling dynamic processes in the cochlea using otoacoustic emissions, 2010.

Vol. 9: *Sylvain Favrot*, A loudspeaker-based room auralization system for auditory research, 2010.

Vol. 10: *Sébastien Santurette*, Neural coding and perception of pitch in the normal and impaired human auditory system, 2011.

Vol. 11: *Iris Arweiler*, Processing of spatial sounds in the impaired auditory system, 2011.

Vol. 12: *Filip Munch Rønne*, Modeling auditory evoked potentials to complex stimuli, 2012.

Vol. 13: *Claus Forup Corlin Jespersgaard*, Listening in adverse conditions: Masking release and effects of hearing loss, 2012.

Vol. 14: *Rémi Decorsière*, Spectrogram inversion and potential applications for hearing research, 2013.

Vol. 15: *Søren Jørgensen*, Modeling speech intelligibility based on the signal-to-noise envelope power ration, 2014.

Vol. 16: *Kasper Eskelund*, Electrophysiological assessment of audiovisual integration in speech perception, 2014.

Vol. 17: *Simon Krogholt Christiansen*, The role of temporal coherence in auditory stream segregation, 2014.

Vol. 18: *Márton Marschall*, Capturing and reproducing realistic acoustic scenes for hearing research, 2014.

Vol. 19: *Jasmina Catic*, Human sound externalization in reverberant environments, 2014.

Vol. 20: *Michał Feręczkowski*, Design and evaluation of individualized hearing-aid signal processing and fitting, 2015.

Vol. 21: *Alexandre Chabot-Leclerc*, Computational modeling of speech intelligibility in adverse conditions, 2015.

Vol. 22: *Federica Bianchi*, Pitch representations in the impaired auditory system and implications for music perception, 2016.

Vol. 23: *Johannes Zaar*, Measures and computational models of microscopic speech perception, 2016.

Vol. 24: *Johannes Käsbach*, Characterizing apparent source width perception, 2016.

Vol. 25: *Gusztáv Löcsei*, Lateralized speech perception with normal and impaired hearing, 2016.

Vol. 26: *Suyash Narendra Joshi*, Modelling auditory nerve responses to electrical stimulation, 2017.

Vol. 27: *Henrik Gerd Hassager*, Characterizing perceptual externalization in listeners with normal, impaired and aided-impaired hearing, 2017.

Vol. 28: *Richard Ian McWalter*, Analysis of the auditory system via synthesis of natural sounds, speech and music, 2017.

Vol. 29: *Jens Cubick*, Characterizing the auditory cues for the processing and perception of spatial sounds, 2017.

Vol. 30: *Gerard Encina-Llamas*, Characterizing cochlear hearing impairment using advanced electrophysiological methods, 2017.

Vol. 31: *Christoph Scheidiger*, Assessing speech intelligibility in hearing-impaired listeners, 2018.

Vol. 32: *Alan Wiinberg*, Perceptual effects of non-linear hearing aid amplification strategies, 2018.

Vol. 33: *Thomas Bentsen*, Computational speech segregation inspired by principles of auditory processing, 2018.

Vol. 34: *François Guérit*, Temporal charge interactions in cochlear implant listeners, 2018.

The end.

To be continued…

Cochlear implants (CIs) are a neural prosthesis that allows severely hearing-impaired listeners to understand speech in quiet. However, listening to a single person's voice among many can be challenging for most CI listeners. In such scenarios, sounds from multiple sources compose a complex acoustic waveform. To hear out an individual source (e.g. a speaker), the auditory system needs to perceptually group the sound mixture into auditory streams. However, few studies have investigated the principles that may allow CI listeners to perceptually segregate and selectively listen to individual sound sources. This thesis combined behavioral and electrophysiological measures to address the fundamental questions of whether CI listeners can perceptually segregate and selectively listen to an individual sound source. The results showed that perceptual differences elicited by varying either the place or the pulse rate of the electrical stimulation allowed CI listeners to perceptually segregate and selectively listen to a target stream. Moreover, the effects of selective attention could be measured using electrophysiological recordings of the brain activity. The results also suggested that the abilities of CI listeners to suppress a competing stream and to initially select the stream of interest may be limited, potentially contributing to the challenges which they experience in complex listening scenarios.

DTU Electrical Engineering
Department of Electrical Engineering

Ørsteds Plads
Building 348
DK-2800 Kgs. Lyngby
Denmark
Tel: (+45) 45 25 38 00
Fax: (+45) 45 93 16 34
www.elektro.dtu.dk