Sequential Stream Segregation in Electric Hearing based on Rate Pitch and Interaural Time Differences

Master's Thesis

Lisa Frohmann, BSc

Graz, October 2022

University of Music and Performing Arts Graz

Graz University of Technology

Advisor:
Doz. Dr. Piotr Majdak

Co-Advisors:
Doz. Dr. Bernhard Laback
Dipl.-Ing. Martin Lindenbeck

Assessors:
O.Univ.Prof. Dr. Robert Höldrich
Dr. Martin Hagmüller

Supported by FemTECH

Zusammenfassung

Cochleaimplantate (CI) ermöglichen ertaubten oder hochgradig hörbeeinträchtigten Menschen, ihre akustische Umgebung wieder wahrzunehmen. Derzeit verfügbare CI-Systeme erzielen gute Sprachverständlichkeit in leiser Umgebung, übertragen aber nur unzureichend Information über Tonhöhen und – im Falle einer bilateralen Versorgung – interaurale Zeitdifferenzen (engl.: *interaural time differences, ITDs*). Gerade für das Lokalisieren von Schallquellen sowie die Separation eines Zielsignals von Störquellen (wie Hintergrundgeräusche oder andere Schallquellen) werden die genannten Merkmale jedoch benötigt.
In dieser Masterarbeit wurde die Fähigkeit zu so genannter *Voluntary Sequential Stream Segregation* bilateral CI-versorgter Personen untersucht. Es handelt sich dabei um eine bewusste kognitive Trennung zweier akustischer Signale. Ähnlich wie in der spontanen menschlichen Kommunikation, wurden in den entsprechenden Experimenten Signale nicht gleichzeitig, sondern ihre Teile rasch sequentiell alternierend dargeboten. Die Stimulation einzelner Elektrodenpaare erfolgte dabei direkt über ein CI-Interface.
Pilot- und Hauptversuche wurden unter Anwendung des *Rhythmic Masking Release* Paradigmas durchgeführt. In diesem Paradigma ermöglichte die Variation von Tonhöhen- und ITD-Information zwischen einem Ziel- und Störsignal bei vorhandener Segregationsfähigkeit das Erkennen von rhythmischen Mustern, welche bei identem Ziel- und Störsignal nicht wahrgenommen werden konnten. Um einen optimalen Arbeitspunkt zu erreichen, wurden die Versuchspersonen vor dem Haupttest in der Unterscheidung dieser Rhythmen trainiert.
Die erhaltenen Resultate beleuchten die Rolle der Tonhöhe und ITD im Prozess der Segregation auditorischer Signale. Die gewonnenen Erkenntnisse zeigen Möglichkeiten und Grenzen der CI-Systeme auf und unterstützen die Weiterentwicklung von Stimulationskodierungen für besseres elektrisches Hören in anspruchsvollen Alltagssituationen.

Abstract

For users of cochlear implants (CIs), it can be exhaustingly difficult to focus on a speaker in an acoustically crowded environment because of sub-optimal encoding of the audio signal. Particularly, CI users are deprived of detailed monaural and binaural timing information, i.e., rate pitch and inter-aural time differences (ITDs), respectively, both of which usually support the cognitive task of voluntary stream segregation in a multi-source environment.

In this master's thesis, the *rhythmic masking release* paradigm was employed with bilaterally implanted listeners to study the effects of rate pitch, ITDs, and their combination on voluntary stream segregation in a direct electric stimulation of listeners CIs. Stimuli were presented under well-controlled conditions at a single ITD-matched electrode pair. The cues were applied to the target and distractor streams, potentially enabling the segregation of rhythmic patterns formed by the interleaved target and distractor signals.

The results of this thesis show a significant synergy effect in the contribution of both pitch and ITDs to the ability to segregate streams. This outcome sheds light on the basic perceptual limits in sequential grouping of sound sources with CIs and will guide the future CI development with the ultimate goal to improve speech understanding in noise.

Affidavit

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used. The text document uploaded to TUGRAZonline and KUGonline is identical to the present master's thesis.

Date

Signature

Acknowledgments

I would like to express my deepest gratitude to the many people who helped me finish this thesis and who supported me throughout my studies in the roles of advisors, mentors and friends.

First, I would like to thank my advisor Piotr Majdak for his professional support and advice, offering help in all aspects of this work at any given time and with great humor.

I would like to thank Bernhard Laback for introducing me to the world of CIs through his lectures and my internships at the ARI, sparking my interest to take part in the research he is so passionate about.

I'm deeply thankful for the help of Martin Lindenbeck who offered advice and input throughout the whole project, flooded me with literature and ideas, but always stayed to dive through any difficulties together. It means a lot to me.

I would like to thank everybody at the ARI, especially Michael Mihocic and Katharina Pollack for their help in the preparation of the listening experiments. Furthermore, I would like to thank the CI listeners for their enthusiastic participation and feedback, making the experiments a very enjoyable and instructive part of my experience.

I would also like to thank Robert Höldrich and my fellow audio engineering students in Graz for all the advice I received throughout my many years of studying.

Finally, I would like to express my heartfelt thanks to my family and friends for supporting me through all ups and downs, lending me their ears and giving advice or a hand in any situation. I deeply appreciate having a village of wonderful people to count on.

I'm endlessly thankful to Thomas for his patience, serenity and unconditional support.

Contents

1 Introduction

When confronted with hearing loss, the patients' quality of life is impaired as they become limited in their actions and interactions on a daily basis. While in some cases the usage of hearing aids can help, there are severe cases where dysfunction occurs in the inner ear and the auditory nerve can no longer receive the needed stimulation. Fortunately, cochlear implants (CIs) offer a way to regain the sense of hearing. The development of this neuroprosthesis is a story of success written over the past 65 years (Wilson and Dorman 2008), bringing back the ability to understand speech and take part in everyday communication.

With all its benefits, electric hearing with CIs is by far no perfect replacement for normal hearing. Research is ongoing to understand the mechanisms of both the normal and the electric hearing, as they differ substantially due to the nature of the stimulation and the way of processing by the brain. With currently used stimulation techniques like continuous interleaved sampling (CIS) (Wilson et al. 1991), CI systems lack to transmit essential details about the acoustic environment. In fact, in CIS, fine structure (FS) of the auditory signal is discarded in favor of a good representation of the signal envelope. Therefore, making use of cues like temporal pitch (especially in the lower frequency region) and precise interaural timing information can be problematic for CI listeners. However, these cues are needed to localize and to discriminate between sources, especially in challenging situations when acoustic information arrives at once or in interleaved streams. Natural human conversation consists of small, heavily competing portions of speech and the brain makes use of various cues to assign signal parts to separate streams. This means that a CI user, if not provided with salient cues, can hardly perform segregation of multiple sound sources unless additional information like loudness differences are present. It is therefore of particular interest to develop stimulation strategies making temporal signal information as accessible as possible.

In this thesis, I focus on a psychoacoustic evaluation of CI users' ability to segregate streams based on temporal cues. As an introduction to this topic, the remaining sections of this chapter describe mechanisms of stream segregation in normal hearing, followed by a general introduction to electric hearing with CIs focussing on temporal pitch and perception of interaural time differences (ITDs), further followed by a description of studies that cover the topic of sequential stream segregation in CI listeners with timing cues, closing with the formulation of the specific thesis goals.

1.1 Stream Segregation

The human auditory system is able to determine the presence of multiple sound streams, even if they appear simultaneously. Research is still ongoing to understand which peripheral pitch and timing cues are needed as a preprocessing mechanism to enable the forming of objects from acoustic features (Bregman 1978) by the higher processing stages in the brain.

Bregman and Campbell (1971) described the phenomenon of primary auditory stream segregation for simple rapid tone sequences. A series of alternating high and low frequency tones was presented to the listener and a perceptual splitting of the series into individual streams of high and low pitch was observed. They found that this subjective assignment to separate groups is influenced by the frequency difference and the rate at wich the tones are presented. The higher the rate, the smaller **frequency differences (FDs)** were needed to enable the perceptual splitting into streams.

Following their findings, Van Noorden (1975) extended the evaluation by including the intent of the listener as a factor to segregation (see also Paredes-Gallardo et al. 2018), differentiating between voluntary and obligatory stream segregation. Voluntary stream segregation means that the listener is actively engaged in the segregation task, i.e., actively following a specific speaker against competing sounds. The process can be described as a top-down, conscious, and attentional way of processing (Stainsby et al. 2011). On the other hand, the obligatory (or involuntary) stream segregation causes sounds to fall into separate streams even if the listener is focused on hearing the sequence as a whole. In experiments with FDs, Van Noorden (1975) assigned specific boundaries to each of the two segregation types: The fission boundary, defined as the FD below which sounds can no longer be segregated, and the temporal coherence boundary, defined as the FD limit above which two streams can no longer be integrated into one. They found the fission boundary to be at smaller FDs than the boundary for temporal coherence.

Auditory streaming is cumulative (Bregman 1978): The auditory system gathers information about the incoming sounds over a period of seconds which finally results in segregation. This so-called build-up is influenced by sudden changes in the perception of a sequence or in the sequence itself which can lead to a fallback into a single fused stream, preparing the auditory system for a possible new income. With FDs as a cue for segregation, build-up becomes faster with larger separation of frequency. Furthermore, the build-up decays exponentially within a few seconds with longer decay for musicians than for non-musicians (Beauvois and Meddis 1997, Moore and Gockel 2002). Changes in attention of the listener by means of focussing on a different ear, spatial location, and a new auditory object can reset the build-up process. All these factors suggest that segregation involves not only peripheral processing of acoustic features but also strong central mechanisms (Moore and Gockel 2002).

While spectral cues provide salient information for segregation, ITDs do not seem to be as important. Stainsby et al. (2011) and Füllgrabe and Moore (2012) both investigated the effect of ITDs on sequential stream segregation in a delay detection task. While Stainsby et al. (2011) used complex tones and observed obligatory streaming ability, Füllgrabe and Moore (2012) tested sinusoidal signals in an obligatory streaming procedure as well as additional subjective segregation judgments. ITDs were tested from 250 to 2000 μs. For complex tones presented between 500 and 707 Hz, ITDs had a significant effect on detection performance for 500 μs and more. However, sensitivity for the ITD with sinusoidal signals was poor overall, especially for ITDs greater than 1000 μs, i.e., beyond the physiological range in normal hearing (NH). They concluded that ITD has only a weak effect on segregation for successive tone burst sequences.

David et al. (2015) investigated the effects of ILDs, ITDs, and coloration (monaural spectral differences) on obligatory segregation using speech shaped noises. Results show that ITD and coloration cues were of greater importance in a temporal discrimination task when compared to ILDs. Reliable and consistent ITD cues across frequencies that lead to a clear lateralization were crucial to the segregation performance. As they used an objective task following a rhythmic discrimination paradigm, an indirect measurement of streaming ability was conducted. To check that they actually tested stream segregation ability, they correlated the data to the results of a subjective task with the same conditions, which confirmed the adequacy of the paradigm.

Vliegen and Oxenham (1999) examined the relationship of spectral and temporal pitch information to sequential stream segregation in a voluntary segregation tasks. They showed that presenting complex tones consisting of high, unresolved harmonics in a melody recognition task can still lead to a segregated percept, indicating that periodicity is sufficient when the spectral information is not available. Based on these results and considering contradictive results in the literature, they formed the general conclusion that stream segregation is possible with sounds differing in either the spectral or the temporal aspects. They emphasized the influence of temporal pitch information on segregation offering opportunities for the design of stimulation algorithms for CIs: If listeners were only supplied with exploitable rate information, segregation in challenging listening environments could be enhanced.

Vliegen et al. (1999) also examined the relationship of spectral and temporal pitch information in an obligatory segregation task, but they discovered some differences in their findings to Vliegen and Oxenham (1999). Their results showed a smaller effect for periodicity than for spectral differences, which they interpreted as an effect of the listening situation. They assumed that the temporal pitch has a strong effect when segregating sounds yields an advantage, whereas it is less important when integration is favorable in the current task. They hypothesized that depending on the desired outcome, pitch cues might be weighted differently, namely, if integration is beneficial, differences will be ignored and if segregation is more desirable, listeners will exploit all available cues.

Factors such as age and hearing loss can also cause degradation of the stream segregation ability. As people get older, they might experience difficulties in understanding speech in multi-source situations, caused by deficits in the peripheral auditory system or cognitive processes. Of course, hearing loss is not limited to the elderly, as congenital impairment or loss induced by trauma or diseases also affect people of all ages. Füllgrabe and Moore (2014) studied the ITD-based stream segregation abilities of younger NH, older NH, and older listeners with hearing loss. Surprisingly, their results showed a comparable performance for the younger and older NH, indicating that age is not affecting the mechanism of ITD-promoted segregation, although average sensitivity to ITDs seems to degrade with age. However, in the group of older listeners with mild to moderate sensorineural hearing loss, the ability of ITDs to induce obligatory segregation was strongly reduced.

Middlebrooks and Onsan (2012) conducted an interesting study about stream segregation with high spatial acuity. They evaluated listeners' voluntary segregation abilities due to spatial separation of streams in a rhythm identification task. Their elegant and simple approach of applying an objective testing paradigm called *rhythmic masking release* caught my attention. Their results in NH listeners gave rise to the idea that bilateral CI users may benefit from exploiting salient ITD information – if they only had access to it. Therefore, Section 1.4 of this thesis focuses on the *rhythmic masking release* paradigm and its application in the CI experiments that were conducted for this body of work.

1.2 Electric Hearing with Cochlear Implants

The CI System Electric hearing requires a combination of hard- and software that is stimulating the functioning fibers of the listener's auditory nerve. Fig. 1 shows a basic representation of all parts of the CI system. In general, it consists of

- an external speech processor carrying a microphone and battery pack that is placed directly behind the ear, which is
- connected to a transmitter that transmits power and the processed signal information across the skin to
- the implanted receiver, i.e., the actual implant, which decodes the signal to generate electric signals which are then passed to
- an array of spaced electrodes inserted into the scala tympani of the cochlea, stimulating the remaining endings of the auditory nerve at specific cochlear locations.

There are many currently used processing strategies for CIs like CIS, SPEAK, n-of-m, HiRes or ACE (see Wilson and Dorman 2008 for a broad overview). The CIS strategy is a common processing option in most of the implant systems in use today and produces satisfying results regarding speech understanding in quiet.

Basic CIS processing steps are illustrated in Fig. 2. A bandpass filterbank filters the signal from the microphone into n channels followed by an envelope detector stage and a compression that is a nonlinear mapping of the wide dynamic range

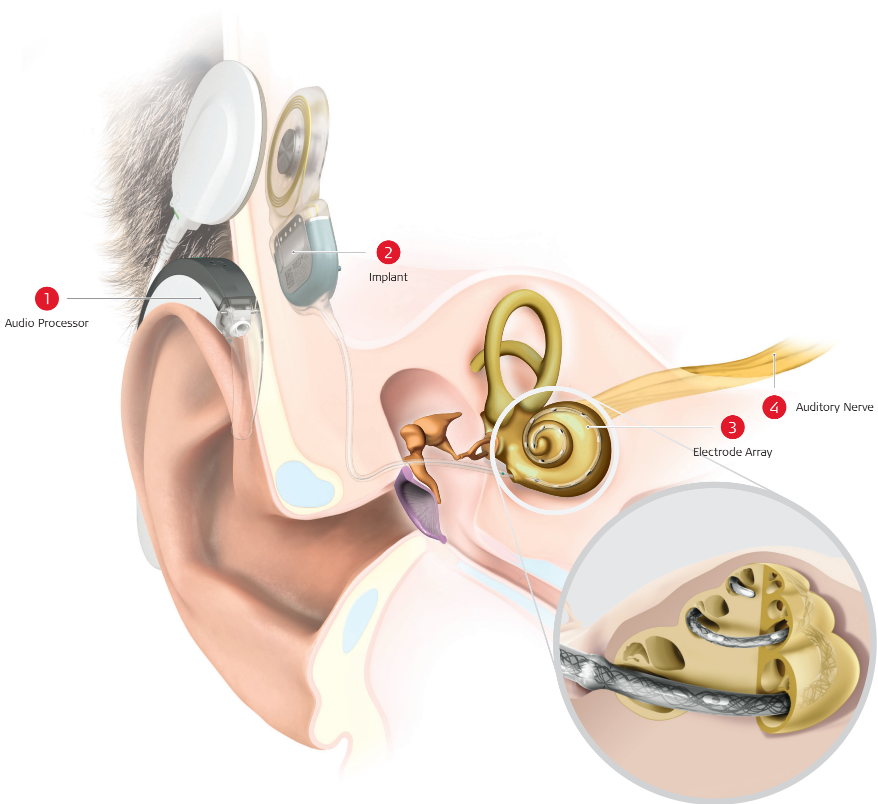


Figure 1: Schematic representation of a CI system (Figure: ©MED-EL).

of the natural acoustic environment (approximately 100 dB) to the very restricted dynamic range of electric hearing of approximately 10 dB. Each of the compressed envelopes is then used to modulate a biphasic electrical pulse train of usually 1000 pulses per second (pps) or more. To avoid interactions of the electric fields of the stimulated electrodes, the pulse trains of the individual channels are temporally interleaved such that the stimulation is not simultaneous.

Pitch Perception Pitch perception in normal hearing is understood as a combination of two mechanisms. On one hand *place theory* (Fearn et al. 1999, Von Békésy and Wever 1960, Von Helmholtz 1954) states that pitch is perceived at the tonotopic position with the maximum vibration of the basilar membrane that is induced by a certain frequency of a sound. Said tonotopy processes high frequencies on the basal and low frequencies on the apical end of the membrane. Furthermore, the pitch scale is mapped nonlinearly from low to high, causing a decrease in frequency selectivity with increasing frequency due to the changing mechanical properties of the cochlear

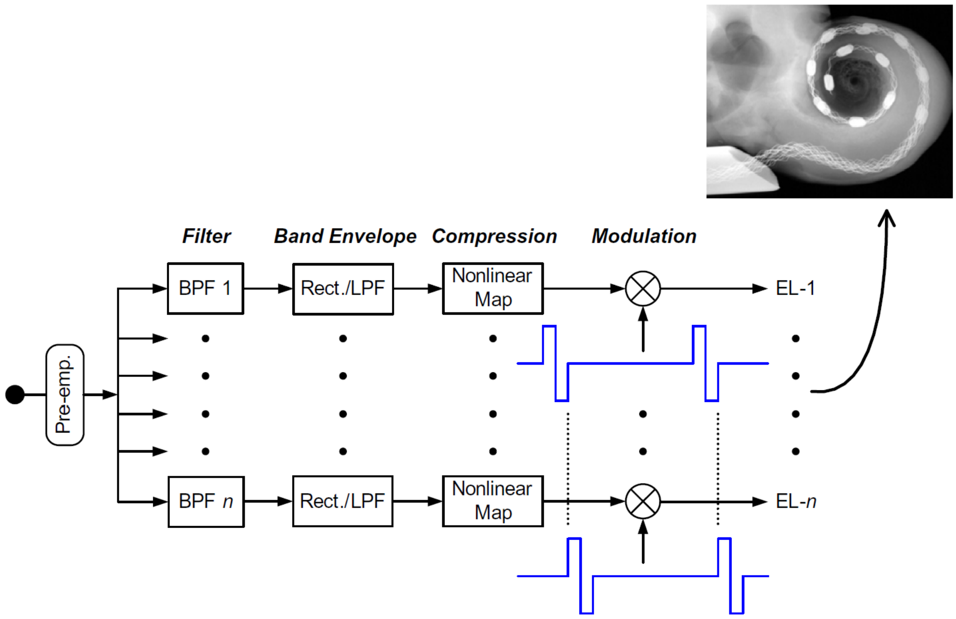


Figure 2: CIS processing strategy (Figure from Wilson and Dorman 2008).

tissues and fluids (Plack 2018, Robles and Ruggero 2001). On the other hand, *rate theory* (Seebeck 1844) suggests that the nerves fire at the rate of incoming sound. The so-called phase locking is a property of the neural system to fire in synchrony with the phase of the currently heard waveform, even above the maximum physiological firing rate of individual neurons. Because phase locking seems to decrease above 1 kHz, not every period causes a firing, but some neuron fires at least at integer multiples of the period (see Plack 2018, Chap. 4.4.4). While neither place nor rate theory alone can explain pitch perception entirely, it is assumed that *rate theory* holds for frequencies up to 1.5 kHz and *place theory* accounts for frequencies greater than 5 kHz, but exact limits are unknown. For the frequencies in between, both mechanisms seem to be used by the brain to form the pitch percept (see a current discussion on phase locking limits in Verschooten et al. 2019).

In CI listeners, pitch resolution is very coarse compared to that of the NH listeners. Common CI systems are limited to a finite number of 12 to 24 frequency channels to which frequency-dependend envelope signal information is assigned. In addition to the poor frequency resolution, most of the temporal fine structure is lost in CI's signal processing and thus the available temporal pitch information is reduced. A compact overview on studies of general rate and place pitch perception in CI listeners can be found in Lindenbeck (2018). They report that CI users perceive temporal pitch only up to approximately 300 pps due to inherent neural limitations (Ihlefeld et al. 2015, Shannon 1983, Townshend et al. 1987). Note, however, that Kong et al. (2009) documented temporal pitch perception even up to 500 pps for some of their participants. Furthermore, the temporal pitch information is solely present in the amplitude modulation of the carrier, hence a reliable pitch perception can be only guaranteed when the modulation frequency is sufficiently oversampled by the carrier (Wilson et al. 1997). While place and rate pitch perception are inevitably connected in healthy cochleas of NH listeners, Moore and Carlyon (2005) found that these two mechanisms can be understood as independent in CI listeners, yielding a possible disadvantage for pitch perception. On the other hand, this offers an opportunity as it enables the studying of the processing of each pitch cue in isolation.

ITD Perception While the perception of pitch is a monaural ability, two ears are needed to determine the location of a sound source based on ITDs and interaural loudness differences (ILDs). While ILDs are prominent for high frequencies due to the acoustic head shadow, ITDs are an important cue in the lower frequency range (Blauert 1997, Laback et al. 2015). Kan and Litovsky (2015) present an extensive review on binaural hearing with CIs, focussing on the limitations that arise from both the device itself and the individual pathology of the users' auditory system. As bilateral CI systems are basically using two separate monaural inputs sampled by independent sampling clocks, ITDs are suboptimally encoded. In addition, microphone positioning and independent automatic gain control can impair the ILDs, confronting the CI listener with unreliable interaural cues.

Laback et al. (2015) reviewed findings about the general ITD sensitivity of CI users tested via research interfaces stimulating a single electrode pair, where listeners were supplied with ITD information in the fine structure ($\mathrm{ITD}_{\mathrm{FS}}$) or via the signal envelope. ITD sensitivity in left-vs-right discrimination tasks showed a high individual variation, from thresholds comparable to those of NH to thresholds being far beyond the ranges of natural ITDs (approx. 800 μs maximum). Fig. 3 shows ITD thresholds collected from 14 studies. $\mathrm{ITD}_{\mathrm{FS}}$ discrimination performance degraded with increasing pulse rates, especially at rates above 400 pps, with some exceptional listeners performing well up to 1000 pps when exploiting both fine structure and onset ITD cues (Ihlefeld et al. 2015, Laback et al. 2015).

While in normal hearing, the overall place of stimulation plays a role for ITD sen-

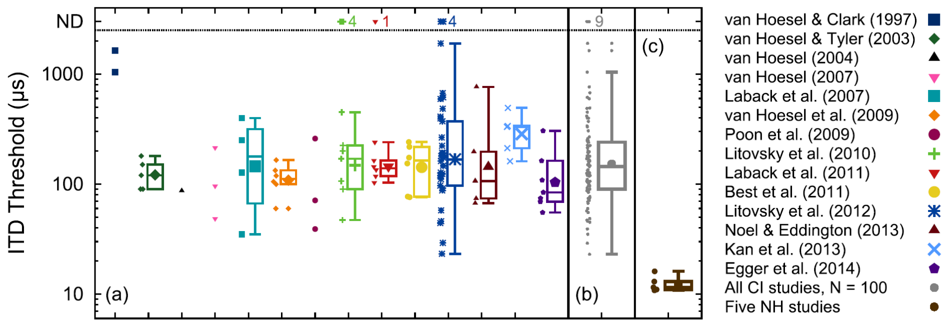


Figure 3: ITD thresholds of CI listeners: (a) data from individual CI studies, (b) pooled data of all 14 CI studies and (c) results for NH experiments (Figure from Laback et al. 2015)

sitivity, this does not seem to be the case for CI listeners. However, an interaural match between two electrodes aiming at eliciting a similar pitch percept seems to be beneficial to the ITD sensitivity (Laback et al. 2015, Van Hoesel 2004). Despite this common approach of pitch-matching, recent studies favor the matching of interaural electrodes that yield the maximum sensitivity to ITDs. Investigations suggest that an adaptation to a mismatched input can be achieved over time for place pitch, while the same adaptation is not observed for ITD sensitivity (Bernstein et al. 2021, Hu and Dietz 2015). Nevertheless, the strong influence of listener-specific parameters on the sensitivity to ITD has to be considered additionally to the hard- and software of the CI. This includes the impact of exposure to binaural sensation in the first years of live, regardless of its acoustical or electrical nature (Laback et al. 2015).

1.3 Stream Segregation in Electric Hearing

CI systems have been refined and improved based on the findings from fundamental research showing new viewpoints on perceptual limits and discovering novel opportunities. Over the last decades, the topic of stream segregation in CI users has come to the fore. Many studies were motivated by advantages that sequential stream segregation mechanisms could offer the CI listeners in everyday communication, helping to separate interleaved sources such as interacting speakers. In this chapter, selected studies on stream segregation in electric hearing are briefly reviewed. They present various approaches to investigate obligatory and voluntary stream segregation abilities measured in objective and subjective tasks.

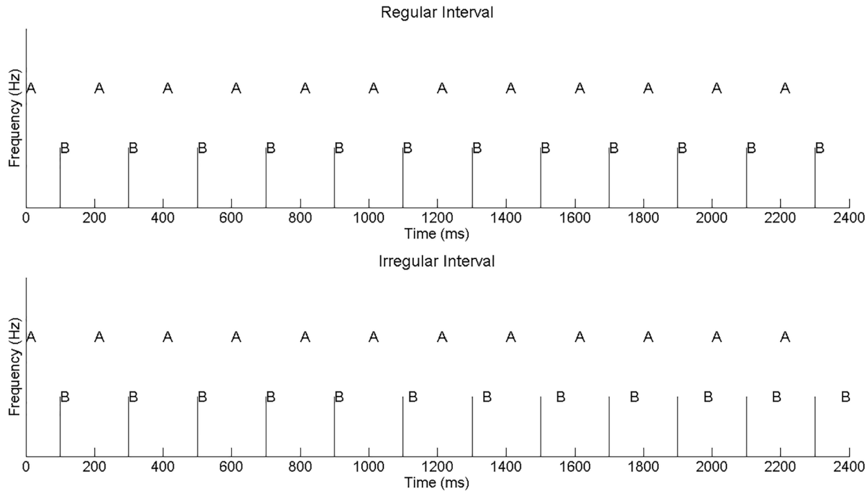


Figure 4: Rhythm sequences from Tejani et al. (2017) with A and B streams differing in the stimulus frequency (acoustic stimulation via speech processor) or the stimulated electrode (direct stimulation) (Figure from Tejani et al. 2017).

Obligatory segregation in an objective task Tejani et al. (2017) examined sequential segregation in NH and CI listeners in an *irregular rhythm detection (IRD)* task. CI users were tested via their own speech processors as well as via direct stimulation over a research interface. Stimulation was monaural and included 60 ms pure tones or narrow-band noises that formed the bursts of a regular and an irregular rhythm sequence (Fig. 4). Listeners were asked which of two intervals contained the irregular rhythm, which was only detectable if A and B streams were integrated, thus segregation hindered a good performance. A and B varied in (center) frequency for acoustic and in electrode position for electric stimulation (place pitch). Additionally, an electrode discrimination task was performed in the CI group, to determine place pitch sensitivity thresholds.

Their results showed a better segregation in NH with increasing FD between A and B. For CI users, tests via speech processor revealed significantly poorer segregation compared to NH as well as CI tests with direct electric stimulation, which was comparable to NH results in some cases. However, they reported a great variability between CI participants. The comparison of the IRD data to the electrode discrimination results showed no significant relation. The authors suggest that discrimination abilities were not reflected in the segregation task.

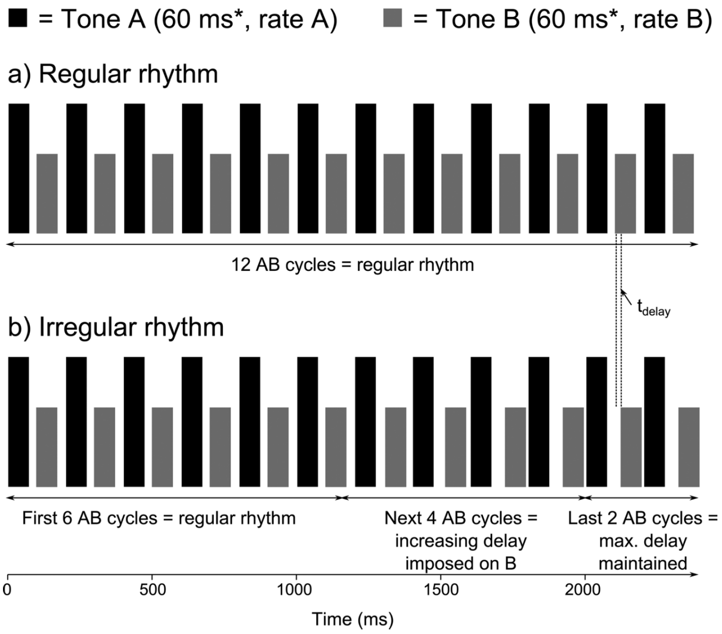


Figure 5: Rhythm sequences from Duran et al. (2012) with A and B streams differing in the pulse rate (Figure from Duran et al. 2012).

In their *rhythm detection* task, Duran et al. (2012) focused on the effect of pulse rate, independent of the place of stimulation. CI listeners were stimulated directly via research interface at one electrode with pulse trains of 60 or 100 ms that form 2.4 or 3.36 s long rhythm sequences, depending on the abilities of the individual listeners. As in the study of Tejani et al. (2017), the interval containing the irregular rhythm should be detected (Fig. 5). While the rate of the A stream bursts was fixed to a base rate of 200, 300, and 800 pps, B stream rates were equal or higher (multiplication of A with Weber fractions of 0, 0.5, and 1).

Results reveal that CI listeners are indeed able to perform segregation based on differences of rate pitch at a constant place of stimulation. Segregation performance decreased with a higher base rate (A), but also increased relatively with a greater Weber fraction (higher rate of B). For some of the listeners they observed evidence for segregation even at a base rate of 800 pps. As this is far beyond the widely assumed limit for rate pitch of 300 pps, they hypothezised that other cues than pitch (i.e. residual loudness differences after the loudness balancing of A to B stimuli) might have been used to solve the given task and hence further investigation is suggested.

Voluntary segregation in a subjective task In most subjective tasks that are concerned with segregation abilities based on pitch cues, listeners are asked if they perceive two seperate streams of differing pitches as opposed to one stream of alternating pitches. In their experiments, Chatterjee et al. (2006) incorporated a *yes/no* task that used electrode place cues in brief sequences (1.2, 1.8, 2.7 or 3.9 s) of A-B-A pulse train triplets (50 ms/burst). In addition, they conducted a discrimination task, where pitch or timbre differences should be detected in a single A-B-A triplet. Lastly, they also measured a general perceptual difference of A and B, asking the listeners for ratings on a simple scale from 0 to 100.

Fig. 6 shows selected streaming, discrimination and rating results for four of five participants. While it can be observed that all listeners could perceive the differences of A and B in the discrimination and difference rating columns, segregation performance was very individual, from asymmetry towards basal electrodes (S1) to a strong build-up effect that comes with stimulus length variations (S5). Additionally, the best performer (S5) was tested at two different locations for A stimuli (Fig. 7a). They were the only subject to concluded experiment 2, repeating the task of segregation and discrimination for modulation of A (fixed, 100 pps) and B (varying) at the same electrode pair for A and B, i.e. segregating the streams based on rate pitch differences (Fig. 7b). Overall, the obtained data indicate that voluntary segregation is possible in the CI-supported auditory system. For S5, rate information via modulation resulted in a streaming percept which was more likely to occur for increasing modulation depth. However, the authors underline that their subjective experimental method might have caused an issue: rather than segregating the streams, listeners could have attended more to pitch differences per se to complete the task. Following this observation that is crucial to the interpretation of the outcome, they propose the use of more objective methods for future experiments to avoid any ambiguities.

Voluntary segregation in an objective task In the study of Cooper and Roberts (2009), segregation abilities were examined in two experiments. An external input socket to the speech processor was utilized for stimulation of the listeners' CIs. While the first experiment focused on obligatory segregation, the second aimed at voluntary segregation that is needed to succeed in an *interleaved melody* task. Fig. 8 displays two examples that each consist of a simple target melody (dark squares) that

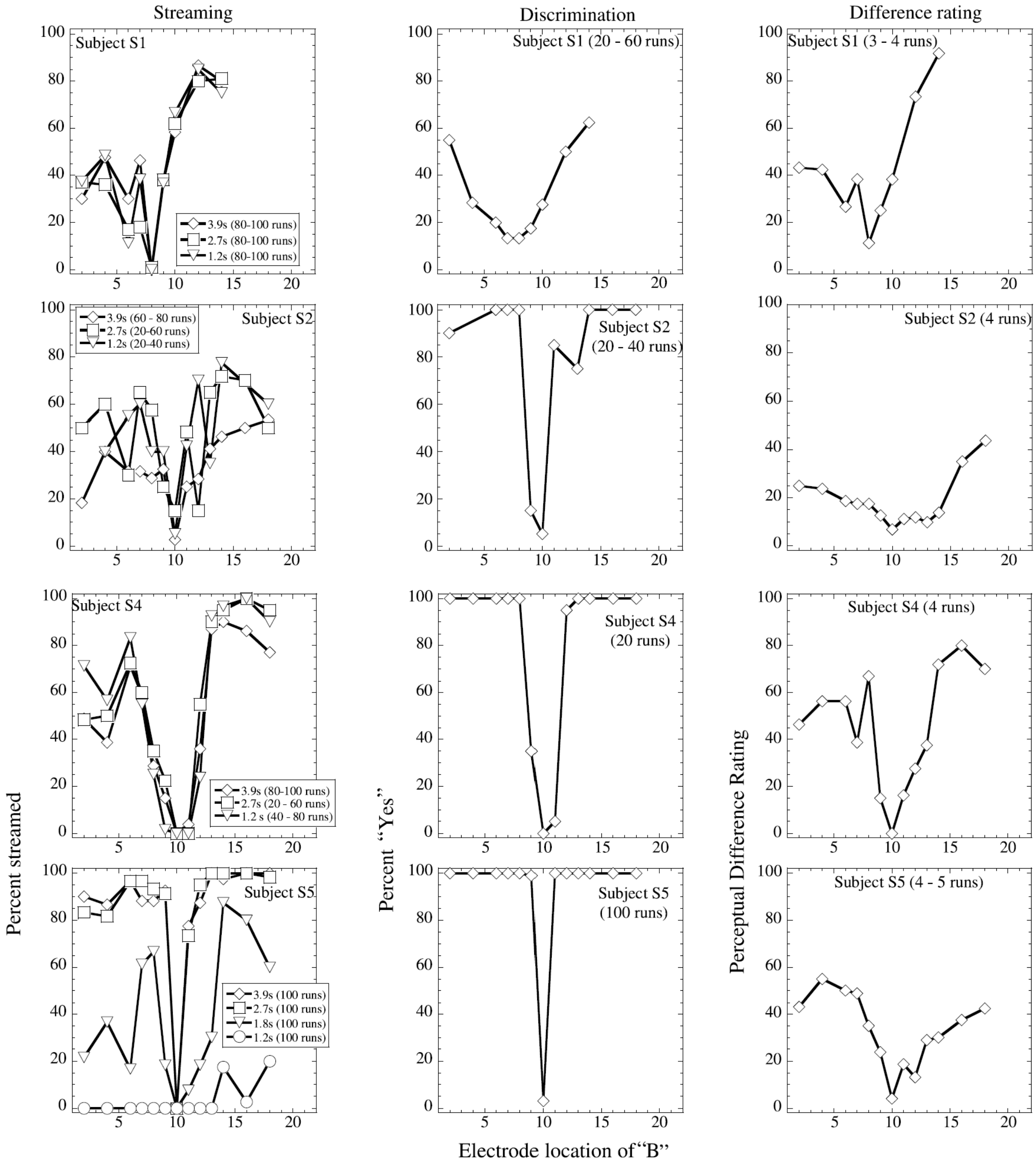
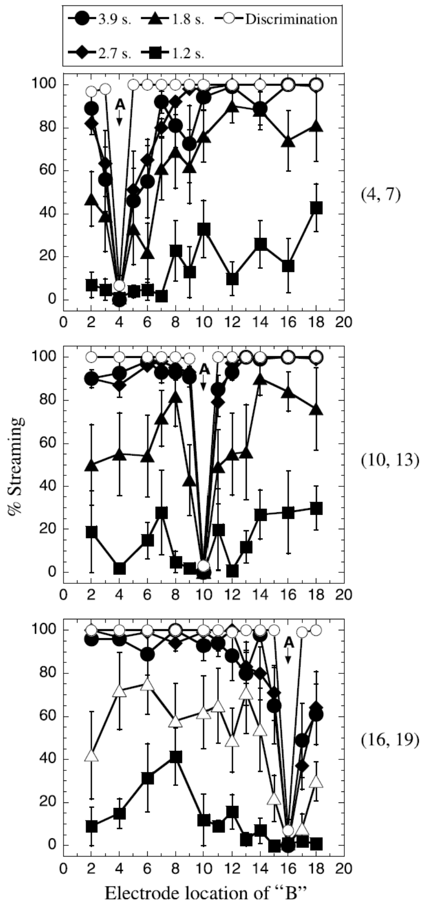
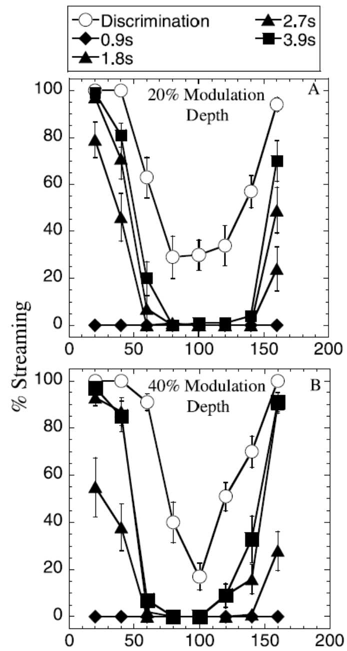


Figure 6: Streaming, discrimination, and difference rating results of four listeners tested on the perception of place pitch cues (figure taken from Chatterjee et al. 2006, modified).



(a) Results for place pitch cues.



(b) Results for rate pitch cues.

Figure 7: Results of S5 from Chatterjee et al. (2006). Place pitch cues were tested at basal (4,7), middle (10,13) and apical (16,19) electrode pairs (Figures from Chatterjee et al. 2006).

is interleaved with a distractor sequence (light squares) in either the same electrode region (middle) or in a more distant region (basal). Melody tones had a duration of 60 ms including on- and offset ramps. In a single interval task, listeners had to decide which of the two target melodies they heard, when the distractor distance varied from middle to basal location. Furthermore, the loudness level of the distractor was altered with respect to the target. The performance was widely consistent over all participants and revealed that identification of target melody was not possible when it was of the same loudness as the distracting melody. Listeners were not able to make use of the presented pitch differences to establish a streaming percept that would allow them to ignore the distractor and focus on the desired melody. Given

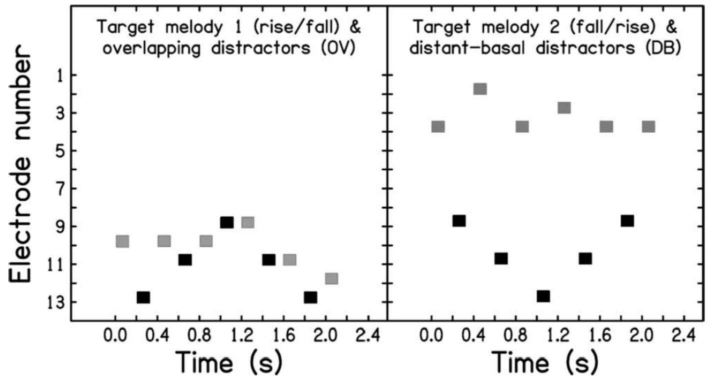
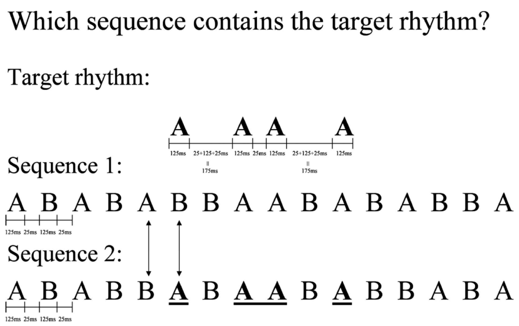


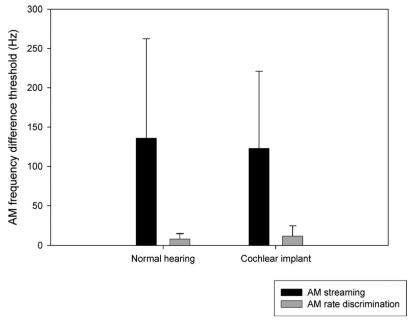
Figure 8: Configuration examples of the *Interleaved melody* task from Cooper and Roberts (2009) utilizing place pitch cues (Figure from Cooper and Roberts 2009).

that the CI users had a clear pitch percept and could discriminate the melodies without the distractors, the results were surprising to the experimenters. They concluded that differing place cues were not effectively enabling neither obligatory (experiment 1) nor voluntary (experiment 2) segregation in the used experimental setup, expressing doubt towards the ability of CI listeners to group incoming sounds in general.

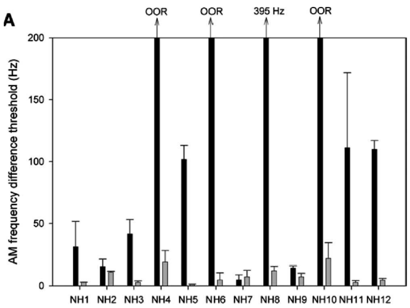
In the experiments of Hong and Turner (2006), all participants used their own clinical processors, i.e. the percept of acoustic stimulation relied on the individual selected stimulation strategy. Streaming experiments were based on frequency differences of pure tones. Widespread results indicated a dependency of stream segregation on place cues across different regions of the cochlea. The performance of some CI listeners was close to NH, while others showed very poor segregation. Assuming that the central processing for streaming is similar across all stimulus frequencies, the authors hypothesized that the variation of performance also reflects peripheral differences, like greater or lesser nerve survival at individual electrodes. Furthermore, they correlated streaming abilities of CI users to performance in speech-in-noise and speech-in-multi-talker-babble tasks, revealing that the CI users with better streaming skills showed better speech recognition as well. However, the researchers noted that remaining loudness cues caused by the acoustic stimulation setup may have lead to a slight but negligible overestimation of segregation abilities. It could further not be ruled out that residual temporal pitch information was available to some listeners



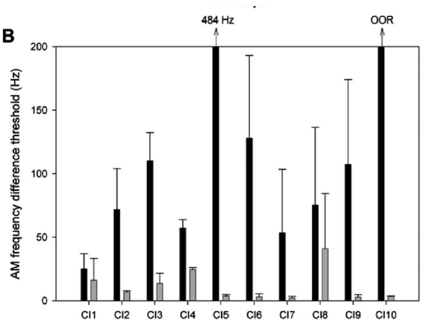
(a) Instructions



(b) FD thresholds for NH and CI listeners



(c) Individual NH thresholds



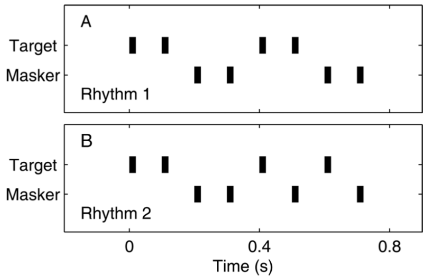
(d) Individual CI thresholds

Figure 9: Task instructions and results from Hong and Turner (2009) (Figures from Hong and Turner 2009).

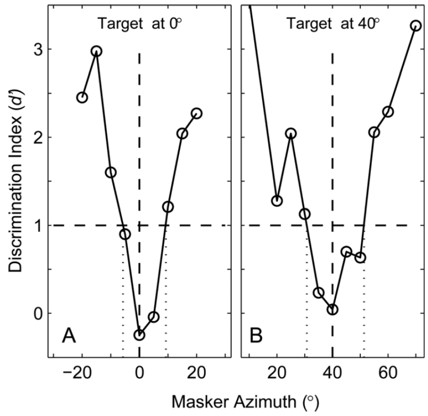
in the lowest stimulation frequency because of individual processor settings.

In Hong and Turner (2009), evidence of stream segregation explicitly using periodicity cues was investigated. Acoustic stimulation via the listener's speech processor was used with sequences of alternating AM noises in an *interleaved rhythmic discrimination* task. The participants were presented with three intervals: the first contained a target rhythm, followed by two sequences of alternating broadband noise bursts that are amplitude modulated (Fig. 9a). Modulation rates for A are fixed (80, 200 and 300 Hz) while B rates varied in increasing logarithmic steps. If the perceptual difference between A and B was large enough and streaming occurred, the subjects could find the target rhythm in sequence 1 or 2. By deliberate design of the rhythmic patterns, focussing on a particular stream A or B was not required as either contained the target rhythm. A and B bursts were played back with a loudness rove of 3 dB to blur any unwanted loudness cues. An AM rate discrimination task was conducted in addition to the streaming task, disclosing sensitivity thresholds. Individual and group results are shown in Fig. 9, indicating that the performance of pure rate discrimination was significantly better than the streaming ability for both NH and CI listeners. Data from individual subjects offer additional insight and lay open astonishing differences within each group, from NH with extremely poor to exceptionally good streaming performance and CI users performing both comparably and worse than the former. It was concluded that the majority of subjects could use temporal pitch cues for segregation and the general ability of discrimination presented no limiting factor. The authors moreover determined modulation depth thresholds needed for streaming and detection at presented base rates. The data suggest that these AM cues are more important for lower rates (80 Hz) than for the perceptional ceiling rate of 300 Hz. Therefore, in a real life scenario demanding the segregation of male and female voices, temporal envelope cues could be more useful concerning lower voice F0s. In their summary the authors point out two limitations of current CI stimulation that are, however, crucial to streaming ability: first, some CIs are simply not transmitting the cues, because the cut-off the low-pass for temporal envelope is set to rather low rates. Second, if cues are transmitted, they are often not sufficiently oversampled by the carrier to ensure reliable representation. Consequently, Hong and Turner (2009) strongly encourage the incorporation of temporal cues in future CI stimulation strategies.

A more recent study by Paredes-Gallardo et al. (2018) investigated segregation abilities based on rate-pitch cues and the build-up with sequence length. In this case, CI listeners were directly stimulated via a research interface, which guaranteed exact control of the stimulus parameter. A *delay detection* task required the listeners to judge timing between 50 ms bursts of a target stream B in the presence of a temporally irregular distractor stream A. Delay detection got easier, if target (300 pps) and distractor (80, 140, 200, and 260 pps) could be segregated. Sequences were 1.24 or 3.96 s long. Sensitivity scores (d') for delay detection improved for increasing rate differences and effects were larger for the longer sequence due to segregation build-up. When comparing the results of a control condition without distractor to the obtained data, the authors found that the CI users could not entirely ignore the distractor stream despite the benefits of a large rate difference and long sequence, which effectively caused an overall slower build-up. In summary, only a few of these experiments were conducted bilaterally. To the author's knowledge, no study has yet investigated the effects of ITDs on stream segregation in CI listeners.



(a) Rhythm sequences 1 and 2.



(b) d' scores for one listener.

Figure 10: Rhythms and results obtained in the RMR task in the azimuth condition from Middlebrooks and Onsan (2012) (Figures from Middlebrooks and Onsan 2012).

1.4 Rhythmic Masking Release (RMR) Paradigm

The experiments of Middlebrooks and Onsan (2012) are particularly interesting. They aimed at isolating the spatial cues used for voluntary stream segregation by NH listeners, using the *rhythmic masking release* paradigm (see Turgeon et al. 2002, Sach and Bailey 2004). Their paradigm requires the discrimination of two basic rhythms formed by noise bursts (in this case for broadband, low-band and high-band noise) in a single-interval two-alternative forced choice (1I-2AFC) task. Each of the rhythmic patterns consists of interleaved target and masker[1] bursts, as displayed in Fig. 10a. Assuming that target and masker are spatially separated in the azimuth, ILD and ITD cues should evoke for a segregated percept of two streams enabling a differentiation between rhythm 1 and 2. In conditions with co-located target and masker, the rhythms collapse into a single burst sequence at 100 ms burst rate. While the target stayed at 0° or 40° in the azimuth, the masker appeared from various loudspeakers locations with a spacing between sources of 2.5° to 5°.

As opposed to Sach and Bailey (2004) who used a random timing of masker pulses, Middlebrooks and Onsan (2012) used a deterministic masker, being essentially a

[1]Note that in the presented study, the term masker is used to describe a distracting stream that is not temporally masking the individual bursts of the target stream due to the non-simultaneous presentation. I use the term distractor in the following chapters when referring to a masker stream.

copy of the target. This allowed the listener to attend to either target or masker stream to solve the task. Bursts were presented every 100 ms and each sequence of eight target and masker bursts was repeated four times to ensure segregation build-up. The experimental design required a training phase without and with maskers for the listeners to remember and identify the patterns. Some of the participants completed a subjective task to report if they heard one continuous or two separate streams for the same conditions as in the RMR procedure to check if they really used segregation in the assignment. The angular differences needed here presented themselves even smaller than for the RMR, which indicated streaming in the RMR tests.
To investigate the spatial acuity of the segregation task to the very limitations of auditory spatial discrimination, the authors further conducted measurements of the minimum audible angle (MAA). In this procedure, two 20 ms noise burst were presented with an onset-to-onset time of 300 ms. Listeners should indicate, if the second burst was to the left or right from the first which fast fixed at 0° or 40° in the azimuth. Exemplary d' for the azimuth experiments for one participant are displayed in Fig. 10b. Listeners showed a high spatial acuity in the task for lowest individual thresholds being smaller than 2°. For detailed results and descriptions of elevation and ILD related experiments, see Chapter IV and V of Middlebrooks and Onsan (2012).

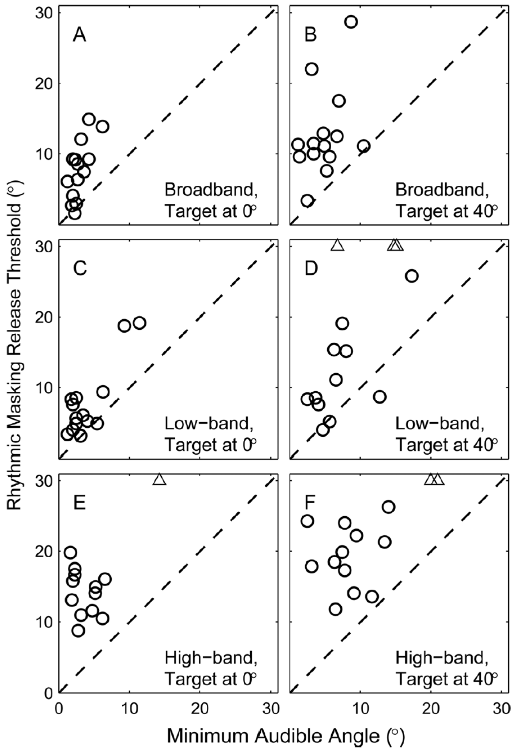


Figure 11: Comparison of RMR thresholds and MAAs from Middlebrooks and Onsan (2012) (Figure from Middlebrooks and Onsan 2012).

Comparison of the results found for RMR and MAA for all listeners and pass-band conditions are shown in Fig. 11. In general, the MAAs were smaller than RMR thresholds. However, some RMR thresholds approached the MAA for the low-band and broadband conditions. The authors give a partial explanation of the performance data by referring to the different onset-to-onset rates of the two tasks (100 ms versus 300 ms) and a possible interfering of the "sluggishness" of the binaural auditory system. The latter describes the poor ability to follow quick variations in binaural cues like ITD, which could have been a problem when presenting rhythms at a 100 ms burst rate. Proposing an additional rationale to the topic, the researchers favor a discrete hypothesis on the processing pathways of the brain: Spatial stream segregation might use different pathways that are representing cues less efficiently than the ones used for the discrimination of source location. Hence, some cues are processed in a way that yields a greater benefit in one task compared to another.

The research of Middlebrooks and Onsan (2012) is not limited to NH, but has explicit implications for CIs. At low frequencies, especially, ITDs could be important cues for segregation which state-of-the-art CIs still lack to transmit. Further experiments on ITD-based segregation abilities in CI users could help to understand possibilities and limitations to improve bilateral CI stimulation.

1.5 Goals and Structure

Based on the trends in stream segregation research, this thesis aimed at utilizing the established and validated *rhythmic masking release* paradigm of Middlebrooks and Onsan (2012) to examine the voluntary sequential stream segregation abilities in CI users. Within this thesis, the investigations were limited to the effects of temporal cues, completing the picture on stream segregation in CI listeners. Cues were presented in the form of rate pitch, ITDs, and a combination of both in the well-controlled environment of direct stimulation. In addition, rate pitch and ITD discrimination thresholds were determined in separate tasks. This enabled a comparison of the RMR performance to the sensitivity thresholds for each cue.

By designing and conducting a listening experiment and subsequent analysis of obtained data, the following research questions were targeted:

- Can CI listeners effectively use rate pitch cues in an RMR task for segregation? If so, what is the effect of the rate pitch?
- Can CI listeners effectively use ITD cues in an RMR task for segregation? If so, what is the effect of the ITDs?

- Assuming a successful streaming in the RMR task, is there a synergy effect of presenting both cues simultaneously?
- Is there a relation between RMR performance and discrimination thresholds?

In this thesis, the RMR paradigm was adapted to consider parameters required in electric hearing and optimized to reduce confounding effects when testing CI listeners. This necessitated the following considerations based on the findings of Middlebrooks and Onsan (2012):

- The target group should consist of bilaterally implanted listeners that are postlingually deafened because they have experienced ITDs in their life before.
- Loudness of all used stimuli should be balanced because ILDs are not of interest in this study and residual loudness differences might confound our results.
- The range of rate pitches and ITDs should facilitate the perception in most CI listeners and should not increase the cognitive difficulty in the task.
- Presentation of ITDs should be optimized by applying them partially to target and distractor streams. Activation of both cerebral hemispheres should reduce direction cues that could be used when the target is fixed at midline.
- Bursts in the rhythm sequences and bursts in the discrimination task should be presented with a similar onset-to-onset duration to render the results of the tasks comparable.

In the following, Chapter 2 describes the implementation and execution of a pilot experiment which evaluated the new tasks on two listeners for an initial set of conditions. Results and listeners' feedback from that pilot experiment formed the base for the main experiment with seven listeners described in Chapter 3. Chapter 4 discusses the results and compares them to the literature. Finally, conclusions from the experiments are summarized in Chapter 5.

2 Pilot Experiment

In the pilot experiment, the overall feasibility of the experimental design and the functionality of the developed testing software were evaluated. After a calibration of the stimuli for each participant (loudness balancing and image centering), the tasks of discrimination and RMR were introduced to the participants, while carefully documenting and evaluating individual feedback to the assignments. All insights gained from the pilot experiment were considered in the tasks and conditions of the main experiment. The following sections describe the methods and procedures used in the different experiment parts and discuss the individual results.

2.1 Methods

2.1.1 Participants

Two subjects participated in the pilot experiment (one female, one male). CI116 had a CI on the right side and was invited for two days to test the experiment parts for rate pitch only. CI24 had a bilateral system and participated in all experiment parts over the course of three days. Both of them received financial compensation for the participation. Detailed subject information is listed in Tab. 1.

	Etiology of hearing loss	Age at testing	Age at deafness onset	CI experience	Implant type[2]		Electrode Nr.	
					L	R	L	R
CI24	progressive	59 yr	40 yr	17 yr	C40+	C40+	6	6
CI116	sudden	58 yr	54 yr	4 yr	-	Synchrony	-	6

Table 1: Information about the participants of the pilot experiment.

[2]Manufactured by MED-EL GmbH of Innsbruck, Austria

2.1.2 Apparatus

Listener's implants were directly stimulated via the Research Interface Box II (RIB2, developed by the Institute of Ion Physics and Applied Physics, University of Innsbruck, Austria) which is offering signal transmission via coils and was connected to a laboratory PC. The listener's own clinical processor is bypassed in this process.

For the new RMR task, the experimental application `RhySeg` was developed using the software framework `ExpSuite`[3] for acoustic and electric stimulation (Mihocic et al. 2012, Acoustics Research Institute (ARI) of the Austrian Academy of Sciences, Austria). It manages the generation of stimuli via MATLAB®, offers an experiment GUI, includes a safe playback of direct electric stimuli to the listener and saves the obtained data for postprocessing. The response input is given by the listener via buttons of a Wingman™ gamepad, reducing additional cognitive load during the experiment due to intuitive and ergonomic handling of the controller. For the loudness balancing of stimuli and for pitch and ITD discrimination tasks the `ExpSuite` applications `DTPitchSIPI` and `DTShortIPI` were used and adapted to handle study-specific input. These applications were developed at the ARI by Martin Lindenbeck (Lindenbeck 2018). Applications support the Wingman™ gamepad and a standard keyboard as user interfaces.

2.1.3 Tasks and Stimuli

The listeners were presented with the following three 2I-2AFC tasks: Either, they had to decide if the stimulus of the second interval was louder or softer (loudness balancing), had a lower or higher pitch (pitch discrimination) or was to the left or the right side (ITD discrimination) as compared to the stimulus of the first interval. In general, the stimuli of these tasks consisted of two single pulse bursts of the same length with defined onset-to-onset timing.

The newly introduced RMR test, however, was a 1I-2AFC task. The participants heard one interval containing one of two rhythm sequences named A and B that were introduced to the listeners in an instruction and training phase. A schematic representation of the rhythms is shown in the upper part of Fig. 12. Each rhythm consisted of target (T) and distractor (D) bursts of pulses. Single bursts had a fixed

[3] https://www.oeaw.ac.at/isf/expsuite

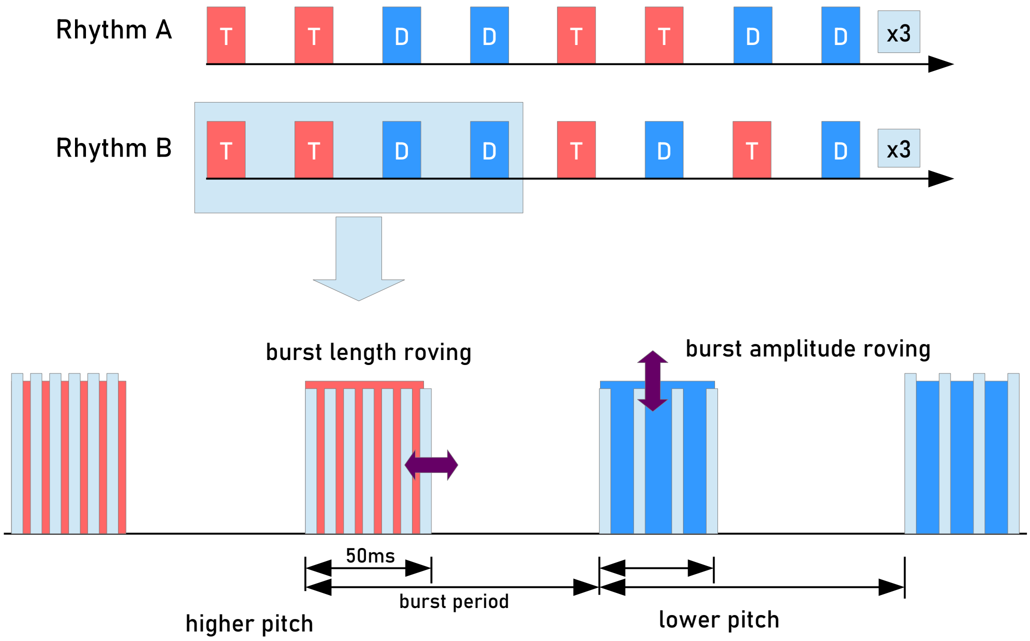


Figure 12: Rhythm sequences A and B (schematic), consisting of temporally interleaved target (T) and distractor (D) streams. The implementation of temporal pitch differences including length and amplitude roving is visualized for a low-rate stimulus.

duration of 50 ms with gaps in between such that bursts appeared at a desired burst period, i.e., playback rate, of at least 100 ms. A single rhythm sequence consisted of four target and four distractor bursts and was repeated three times consecutively, leading to a total number of 24 bursts such that listeners could sufficiently grasp the pattern through segregation build-up. While the rhythms were the same in the first half, target and distractor alternated rapidly in rhythm B while A remained unchanged. As long as neither pitch nor ITDs differed between target and distractor, the two rhythms were identical. Temporal pitch of target and distractor was incorporated by varying the rate (which is equivalent to varying the period) of the pulse train forming the burst. Pilot test conditions were implemented for low-rate (LR) and high-rate (HR) stimuli. For the latter, a carrier pulse train of 2000 pps (500 μs) is modulated with the same desired rates that are used in the LR condition. While no on- and offset ramps are used for LR, the first half of the first AM period and second half of the last AM period at full modulation depth served as a smooth pseudo-ramp. This should avoid a sudden and unwanted cue to the auditory system which is very sensitive to binaural information at the signal onset (Laback et al.

2015, Litovsky et al. 1999). Modulation depth at the steady part of the stimulus was set to 0.3.

The desired burst length of 50 ms of the RMR did not always fit an integer amount of pulses in each pitch period. Therefore, a burst length roving was applied such that bursts had a pulse randomly added or not and therefore occasionally exceeded the original burst length. In this way, length cues associated to specific pitch rates could be avoided. Although loudness balancing of the stimuli was conducted to eliminate loudness cues between conditions, a random burst amplitude roving was introduced to blur any remaining cues. The amplitude of each burst of a sequence was altered by ±3% of the dynamic range while keeping a binaural synchronization to preserve the centered image calibration (introduced in Sec. 2.2.2). Note that the burst amplitude roving was omitted for ITD conditions that are presented without any FD variations. Schematic representations of the roving in amplitude and length can be found in the lower part of Fig. 12.

Besides temporal pitch, target and distractor streams of the rhythm sequences should differ in perceived direction. ITDs between the bursts of target and distractor were incorporated such that the target is not always centered at midline, but desired ITD amounts were split equally to both target and distractor streams by delaying the right and left ear signals accordingly (see Fig. 13 displaying the ITD split for an ITD of 100 μs). This lead to the streams being perceived from slightly to the left for one and slightly to the right for the other stream, depending on the current condition and the size of the ITD.

2.1.4 Conditions

To determine adequate stimuli conditions for the discrimination tasks and the newly introduced RMR task, a variety of initial configurations were prepared. Similar conditions for discrimination and RMR testing were favored to make sure all results could be meaningfully interpreted and compared.

For the fitting and discrimination tasks, *short* (50 ms) and *long* (600 ms) bursts could be selected. *Short* corresponded to the duration of a single burst in the RMR rhythms, while *long* was equivalent to the duration of all target or distractor bursts of one rhythm sequence (12×50 ms). CI116 was tested at LR and HR conditions with short and long stimuli. Due to time restrictions, CI24 was tested solely at LR

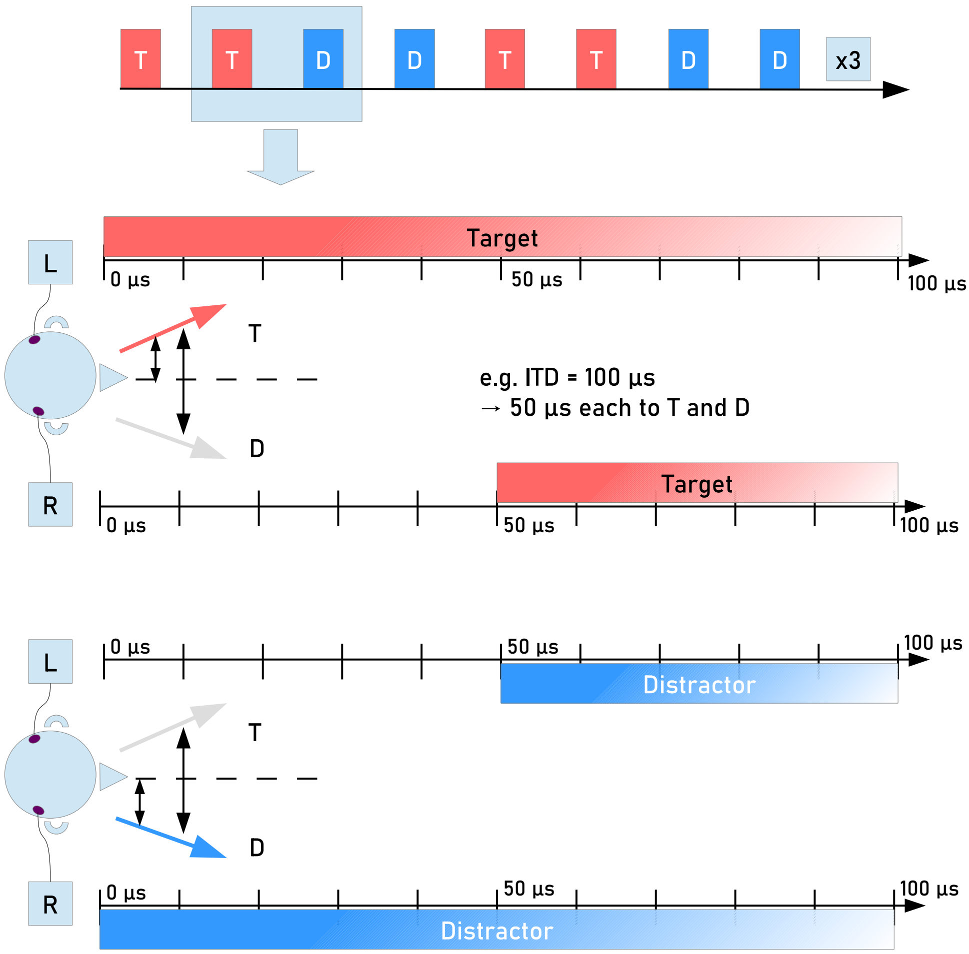


Figure 13: Split of ITDs to target (T) and distractor (D) burst onsets: In this example, the target appears to the left and the distractor to the right due to delay of the opposite ear signal.

conditions with short stimuli. For a burst period of 100 ms the entire RMR rhythm sequence lasted for 2.4 secs and 4.8 secs for a burst period of 200 ms, respectively.

To target a suitable range for FDs, a nominal rate $\bar{R}$ of 118 pps (8.5 ms period) was defined that rate pairs should be closely geometrically centered around: The corresponding lower and upper rates R_{lower} and R_{upper} were then selected to fulfill

$$\bar{R} = \sqrt{R_{\text{lower}} \cdot R_{\text{upper}}} \tag{2.1.1}$$

All rate conditions had periods ensuring a sufficient sampling by the HR carrier. In addition, the FD steps between the pairs were selected to increase in a logarithmic fashion allowing an approximation with musical intervals. Table 2 displays detailed

Rate-pitch pair (pps)	Avg. rate (pps)	Period (ms)	rounded FD (%)	Interval/FD approximation
118 / 118	118	8.5 / 8.5	0	unison
111 / 118	114	9 / 8.5	6	semitone, 6%
111 / 125	118	9 / 8	13	whole tone, 12%
105 / 133	118	9.5 / 7.5	27	2 × whole tone (33%), major third
91 / 143	116	11 / 7	57	4 × whole tone (67%), minor sixth
83 / 168	118	12 / 6	100	octave, 100 %
83 / 200	129	12 / 5	140	octave + 6 semitones (141%) octave + tritone

Table 2: Rate-pitch pairs. Average rates were selected to be closest to nominal rate $\bar{R} = 118$ pps with FDs approximating musical intervals.

information about all pairs and approximations (e.g. interval calculation for the semitone as $2^{\frac{1}{12}} = 1.059$ represents a rounded FD of 6%). The pair with an FD of 140% was used for training purposes and in the discrimination task, but was not included in the RMR main task. The largest pitch period of the selection was set to 12 ms (83 pps). For LR stimuli, it could be ensured that a sound burst of 50 ms consisted of at least four singular pulses. This presented enough information to the listener such that individual pulses would not interfere with the rhythm perception when perceived as separate instances. The upper testing limit was at a 6-ms period (168 pps) which was an octave above the lower bound and still within the limits of CI rate pitch perception ceasing at approximately 300 pps. This was also the case for the highest rate for training purposes that was set to 200 pps, i.e., a period of 5 ms.

ITDs tested in CI experiments are usually larger than the naturally occurring ITDs, i.e., beyond 800 μs, because most CI users have a weak access to ITDs via their clinical processors resulting in a lack of sensitivity thereof. However, ITDs need to be smaller than approximately one quarter of the smallest stimulus period (Majdak et al. 2006). This should avoid the so-called phase ambiguity, which describes the difficulty for the brain to tell if the left or the right side signal comes before the other when ITDs are close to half of a stimulus period (see Hartmann and Macaulay (2014) and Majdak et al. (2006) for more details about phase ambiguity). For the stimuli with rate of 168 pps (6-ms period) in this experiment, the maximal ITDs were thus 1500 μs and 1250 μs for the training stimulus of rate 200 pps (5-ms period). The selected ITDs for all tasks were 0, 100, 400, 800, or 1600 μs. The ITD of 1600 μs was included as the listeners' sensitivity was of interest and no such results were available.

2.1.5 Fitting

Both subjects were experienced listeners and had participated in previous studies of pitch perception at the ARI. The selection of a single electrode pair for this experiment was influenced by previous choices to ensure reliable performance and reduce the time costs of a selection procedure. Details are listed in the participant information table (Tab. 1).

The listeners' overall thresholds (THR) and maximum comfort levels (MCL) for stimulation were determined in their specific stimulation range for current units (cu) via the `ExpSuite Fitt4Fun` fitting software. Phase duration of the biphasic stimulation pulses could be adjusted individually to achieve a sufficient width of the dynamic range (DR) that is needed throughout the testing process. The used fitting signal of either 50 or 600 ms was presented at the nominal rate or AM rate of 118 pps (8.5 ms period) for LR and HR condition, respectively. Comfortable listening levels (CLs) were determined and served as the reference levels in the loudness balancing task.

2.2 Procedures

The following sections explain the procedures used in the preliminary loudness balancing and image centering for all pitch rates, as well as the procedures of the subsequent discrimination and RMR tasks, including the RMR training.

2.2.1 Loudness Balancing

Since it was not the intention to include loudness cues to the temporal cues under test, all stimuli needed to be balanced in loudness. The balancing pretest was conducted for one ear at a time. Using an adaptive 3-up-1-down staircase procedure (Jesteadt 1980), levels were determined such that each of the nine target signals used throughout the tasks (Tab. 2) was perceived as loud as the reference signal of 118 pps (8.5 ms period) at CL. Trials of up- and down-staircases were interleaved in a

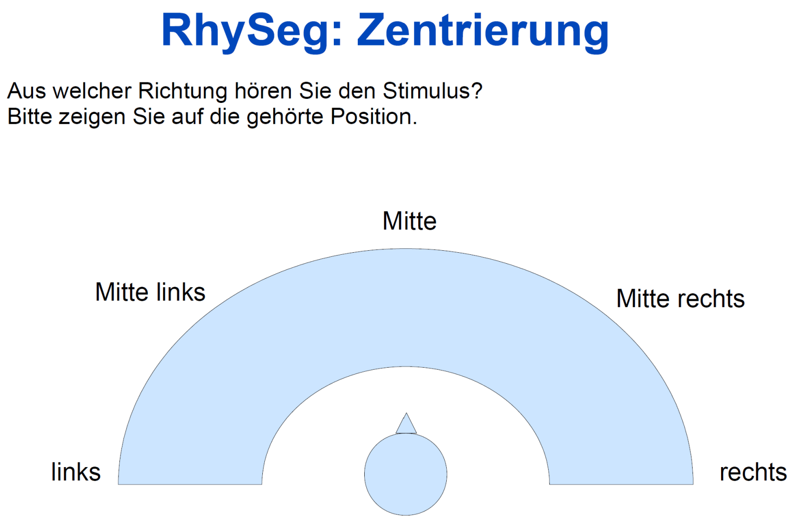


Figure 14: Instructions for the image centering procedure. Listeners were asked to indicate the position of the perceived auditory image to the experimenter.

random fashion. The procedure was stopped after 12 staircase turnarounds and the amplitudes were averaged over the last eight turning points, yielding the balanced amplitude (the reader is referred to Jesteadt 1980 and Frohmann 2022 for detailed information about the balancing procedure in use). The balanced amplitude of a single condition was accepted when the standard deviation (SD) was less than 5 cu. The task was repeated for larger SDs until it reached the acceptable range presuming that listeners performed with better focus in the rerun. Note that aiming for a high measurement precision in the balancing task was excessively extending the duration of this preliminary test.

2.2.2 Image Centering

Balanced amplitudes for each rate condition likely differ between left and right side due to differences in physiology and because the loudness balancing considers each ear separately. This might lead to a shift of the auditory image to the louder side once stimuli are presented bilaterally. To counterbalance this non-centered perception, an image centering procedure was performed. The listener was asked to indicate the position of the auditory image for each pitch period on a chart (Fig. 14). Stimulation amplitudes for each side were manually adjusted by the experimenter via the `ExpSuite` software `LevelDancer` until the listener perceived the sound image as centered. The participant tended to experience minor to major difficulties in the beginning of this task, reporting the auditory image to be inside or behind the head, rather than in front of them, or having it jump rapidly from one side to the other. The listener was advised to take enough time to hear the subtle differences or skip the specific stimulus rate and try again later if there was a strong ambiguity. Despite a possible lack of precision, this procedure offered a simple, intuitive, and time-efficient way to prepare the setup for the upcoming ITD tasks. Furthermore, it gave insight to the individual perception and promoted interaction between the listener and the experimenter which supported a collaborative atmosphere that encouraged the listeners to communicate their sensations without reservations.

2.2.3 Pitch Discrimination

The method of constant stimuli was used in the pitch discrimination task. 100 repetitions per condition were requested for each FD of which equal shares were accounting for a lower-higher (upward order) and higher-lower (downward order) stimulus presentation. All trials were presented in a random order by the application. Depending on the listener's abilities and focus, breaks could be scheduled at a desired percentage of experiment progress.

Results were obtained as percent correct (Pc) scores per condition. However, as this measure does not consider the listener's bias, the so-called sensitivity index d' was calculated with a correction to compensate for response bias due to the presentation order. With showing the The general relation of the probability p and the z-score (Klein 2001) can be implemented in MATLAB® using the inverse error function erfinv(·)

$$z = \Phi^{-1}(p) = \sqrt{2} \cdot \mathrm{erfinv}(2 \cdot p - 1) \tag{2.2.1}$$

which then can be used to calculate the z-scores z_{up} and z_{down} for probability of correct answer p_{c} splitted by up- and downward order as

$$z_{\mathrm{up}} = \sqrt{2} \cdot \mathrm{erfinv}(2 \cdot p_{\mathrm{c,up}} - 1) \quad \text{and} \tag{2.2.2}$$
$$z_{\mathrm{down}} = \sqrt{2} \cdot \mathrm{erfinv}(2 \cdot p_{\mathrm{c,down}} - 1). \tag{2.2.3}$$

This leads to the final d' score through the general connection for d' in a 2-AFC task

(Green et al. 1966):

$$d'_{\text{pitch}} = \sqrt{2} \cdot \frac{z_{\text{up}} + z_{\text{down}}}{2} \tag{2.2.4}$$

For scores of 100% correct, the d' score would approach infinity due to the inverse error function used to compute the z-scores. Therefore, percent correct scores were adjusted to a maximum of 99%, which led to an upper limit $d'_{\text{pitch,max}} = 3.29$.

2.2.4 ITD Discrimination

Similar to the procedure described in the previous section, the ITD discrimination task used the method of constant stimuli with 100 repetitions per condition asked in a random order. In this case, one half of the presentations had a presentation order indicating the stimuli moving left-to-right through the two intervals while the other half was assigned a right-to-left order.

The calculation of the d' was analogous to equation 2.2.4 according to

$$d'_{\text{ITD}} = \sqrt{2} \cdot \frac{z_{\text{to left}} + z_{\text{to right}}}{2} \quad \text{with} \tag{2.2.5}$$

$$z_{\text{to left}} = \sqrt{2} \cdot \operatorname{erfinv}(2 \cdot p_{\text{c, to left}} - 1) \quad \text{and} \tag{2.2.6}$$

$$z_{\text{to right}} = \sqrt{2} \cdot \operatorname{erfinv}(2 \cdot p_{\text{c, to right}} - 1) \tag{2.2.7}$$

representing the z-scores for correct answer probability splitted by movement order. The same limitation of d' to 3.29 is applied.

2.2.5 RMR

The RMR task consisted of three separate parts. Listeners had to complete 100 repetitions per condition for individual pitch and for ITD tasks, whereas there were only 96 repetitions per condition in the pitch-ITD combination due to the combination of the two presentation orders. Results were obtained as Pc and converted to d' scores. The RMR task used a variation of the yes/no method, where the presence of rhythm A was of primary interest. A *correction for response bias* was applied as in Klein (2001) and Middlebrooks and Onsan (2012). The corrected d' was obtained from the z-scores of the hit $z(\text{H})$ (rhythm A played and rhythm A detected) and false alarm rates $z(\text{FA})$ (rhythm B played and rhythm A detected). Following Eq. 2.2.1,

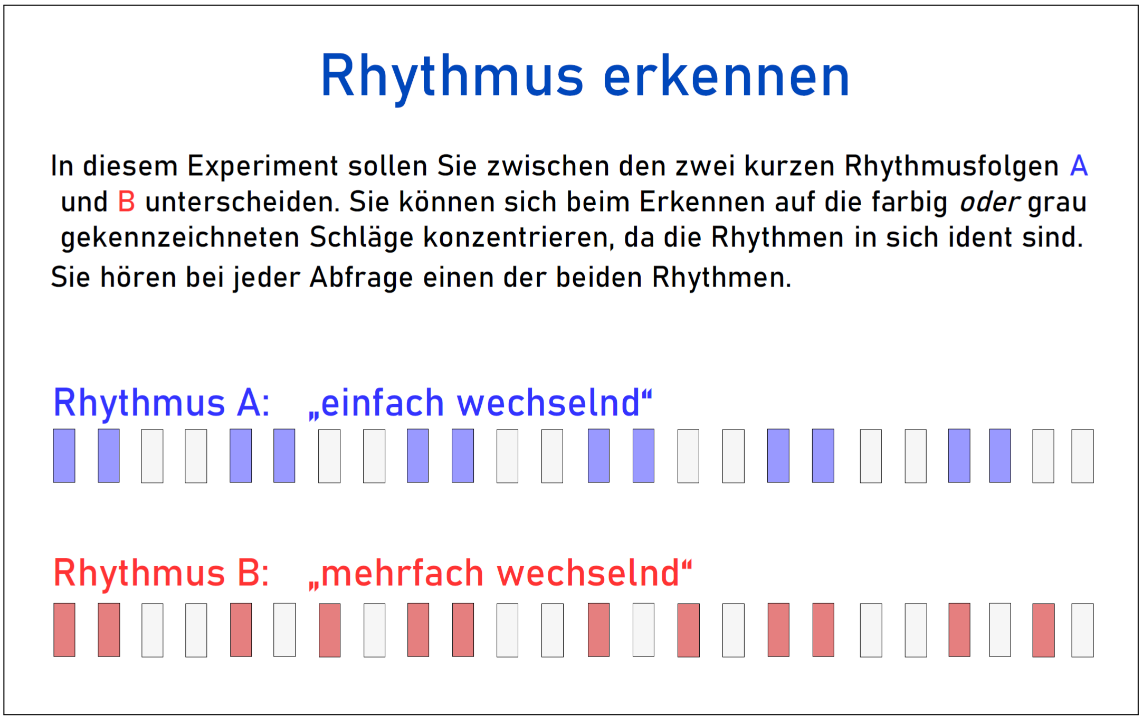
d' was computed as:

$$\begin{aligned} d'_{\mathrm{RMR}} &= z(\mathrm{H}) - z(\mathrm{FA}) && (2.2.8) \\ &= \sqrt{2} \cdot (\mathrm{erfinv}(2 \cdot p(\mathrm{H}) - 1) - \mathrm{erfinv}(2 \cdot p(\mathrm{FA}) - 1)) && (2.2.9) \end{aligned}$$

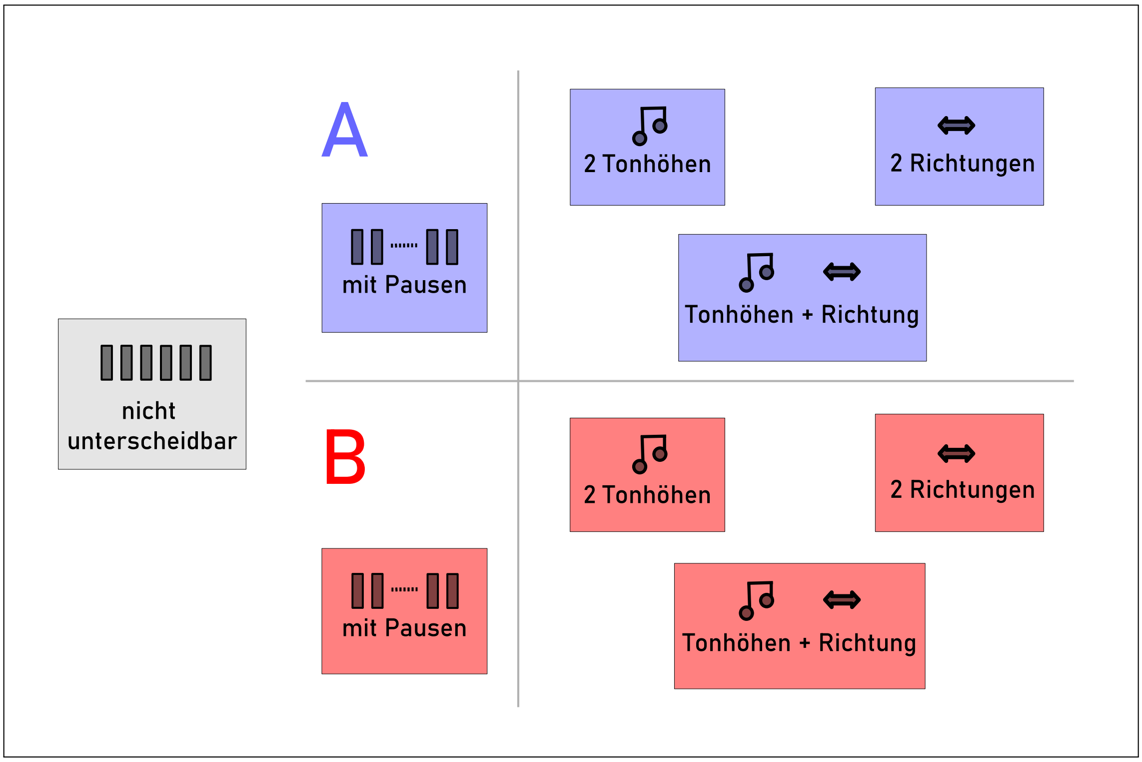
where $p(\mathrm{H})$ and $p(\mathrm{FA})$ denote the probability of hits and false alarms, respectively. As in Secs. 2.2.3 and 2.2.4, the d' score would approach infinity for 100% correct answers and thus the same correction of the maximum was applied. This resulted in the following upper limits of the d' in the RMR tasks with the respective number of repetitions for each parts:

$$\begin{aligned} d'_{\mathrm{RMRmax,100reps}} &= \sqrt{2} \cdot (\mathrm{erfinv}(2 \cdot \frac{49}{50} - 1) - \mathrm{erfinv}(2 \cdot \frac{1}{50} - 1)) && (2.2.10) \\ &= 4.1075 \\ d'_{\mathrm{RMRmax,96reps}} &= \sqrt{2} \cdot (\mathrm{erfinv}(2 \cdot \frac{47}{48} - 1) - \mathrm{erfinv}(2 \cdot \frac{1}{48} - 1)) && (2.2.11) \\ &= 4.0737 \end{aligned}$$

Training The RMR task required a rhythm training because listeners needed to be familiar with the presented rhythms A and B which served as a mental template utilized in the procedure. The rhythm training consisted of an informal and a formal part. First, the task was explained to the listener verbally and in written form using an instruction sheet that illustrates the burst sequences forming the rhythms (Fig. 15a). Subsequently, the singular training conditions were manually selected for stimulation (basic rhythm without distractor stream, FDs of 0% and 140%, ITD values of 2000 and 2400 μs) and the listeners were given time to react to the new sound qualities and memorize the rhythmic patterns. All variations and orders of presentation were explained while the currently heard stimulus type was indicated by hand on the instruction sheet (Fig. 15b). Playback was repeated multiple times by request. Listeners' feedback of any kind was highly encouraged during this stage. Sequences without distractor bursts and uniform sequences without any cues functioned as a sanity check if the listeners could at all perceive and discriminate the basic patterns. Conditions were presented in either 40 or 80 trials per condition.



(a) Introduction to the rhythm sequences A and B.



(b) Instruction for the informal rhythm training.

Figure 15: Instructions for the RMR task.

Balanced and centered amplitudes (cu)											
		Pitch period (ms)									
ID	Amp, stim, range	5	6	7	7.5	8	8.5	9	9.5	11	12
CI116	Bal. LR, short, R3	55	61	64	67	67	67	70	68	72	66
	Bal. LR, long, R2	71	71	72	72	74	74	72	74	74	72
	Bal. HR, short, R3	62	66	63	64	63	63	65	65	64	63
	Bal. HR, long, R2	93	101	91	95	94	93	91	98	99	89
CI24	Bal. L, short, R3	47	50	54	58	55	50	59	54	58	59
	Bal. R, short, R3	64	64	67	72	73	74	74	75	76	78
	Centr. L, short, R3	44	50	54	55	58	53	58	54	60	58
	Centr. R, short, R3	66	67	70	74	73	74	74	73	75	71

Table 3: Balanced amplitudes for CI116 and balanced and centered amplitudes for CI24.

Test As the RMR main tests were performed on the second and third day of testing, the informal training was repeated prior to the main tasks to refresh the participants' memory of the rhythms and conditions. CI116 completed two test sets (LR, HR) for pitch including a 0-condition for 0 % FD. CI24 started out with a diotic RMR pitch task for the originally intended burst period (100 ms) and proceeded with diotic and a left ear tasks at a 200 ms burst period. ITD tests were also conducted at original and half-time burst period. For the combination task, FDs should be selected that yielded poor to mediocre performance in the RMR pitch set. This should leave room for improvement by the hypothesized synergy effect when adding the ITD cue. Overall, a total number of nine combinations of FDs (6, 13, and 27 %) and ITDs (400, 800, and 1600 μs) was tested.

2.3 Results

Table 3 displays the obtained balanced and centered amplitudes in current units (cu) for all stimuli conditions in the selected ranges. Table 4 shows details and results of the formal pilot training in percent correct. Figures 16 to 20 present d' scores for the discrimination and RMR tasks. Note that, as mentioned in Sec. 2.1.4, the two subjects were presented with different conditions.

Training without distractor					
	Burst period (ms)	ITD (μs)	FD (%)	# trials	% correct
CI116	100	0	0	40	100
CI24	100	0	0	40	90
		0	140	40	82.5
		2000	0	80	81.25
Training with distractor					
CI116	100	0	0	40	45
		0	140	40	100
CI24	100	0	0	40	47.5
		0	140	40	57.5
		2000	0	80	51.25
	150	0	0	40	45
		0	140	40	75
		2000	0	80	55
	200	0	0	40	47.5
		0	140	40	80
		2000	0	80	51.25
		2400	0	40	60

Table 4: Training results of CI116 and CI24 without and with distractor.

CI116 Figure 16 shows the d' scores of the pitch discrimination task for CI116 with the dashed line at $d' = 1$ as a criterion for threshold discrimination. LR stimulation encouraged a better overall performance compared to HR, especially for short stimuli. While testing with long stimuli lead to higher scores for greater FDs, the short stimulus facilitated better discrimination already at the smallest FDs. CI116, who classified as musically trained, showed perfect scores throughout the RMR training sets (Tab. 4) and did not report any difficulties with the task. The RMR results are displayed in Fig. 17, indicating a strong discrepancy between the HR and LR conditions in favor of the latter. In both cases a shift in performance can be observed between 27 and 100 % FD. It seems that rhythm identification due to voluntary segregation worked best at larger FDs.

CI24 Fig. 18 shows pitch and ITD discrimination scores for CI24. FDs were tested for an onset-to-onset time of 100 ms between the stimuli of the two intervals, i.e. a 50-ms gap. This was the same gap length as between two bursts of a rhythm

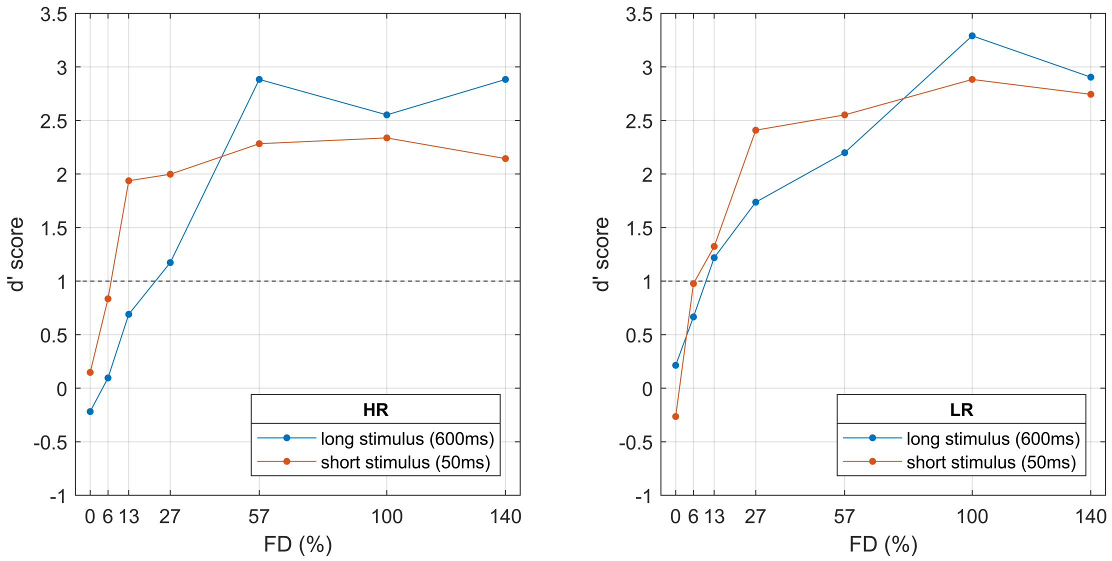


Figure 16: CI116: pitch discrimination d' score for high-rate (HR) and low-rate (LR) stimuli.

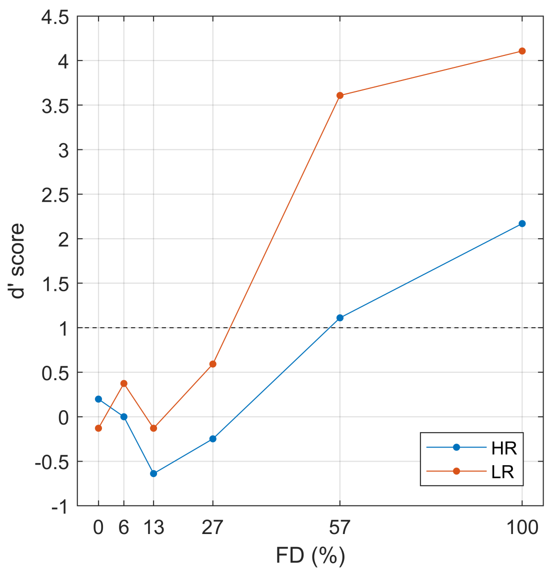


Figure 17: CI116: RMR d' score for pitch condition for HR and LR stimuli.

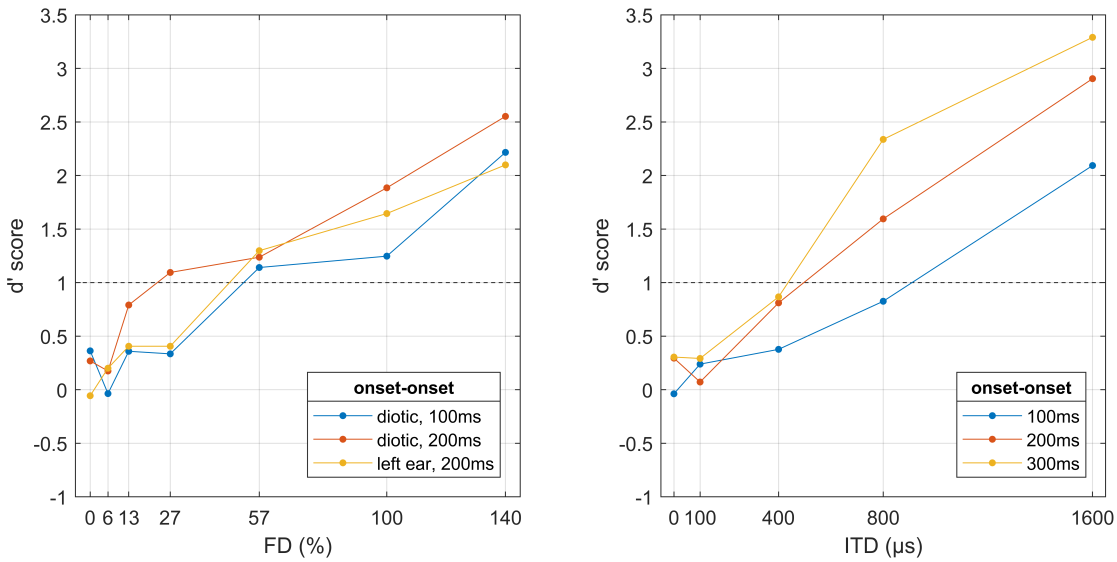


Figure 18: CI24: Discrimination d' scores. Pitch (left) and ITD discrimination (right) results for varying onset-to-onset times of stimuli.

sequence in the RMR task at the original burst rate. Onset-to-onset of 200 ms yielded a 150-ms gap, like in RMR half-time playback. The overall performance at pitch discrimination improved with the larger gap, as long as it was a diotic task. In terms of ITD discrimination, a greater gap had a major beneficial influence on performance for ITDs greater than 400 μs.

In the RMR training, CI24 achieved 80 % correct for the 140 % FD condition at a burst period of 200 ms and 75 % at 150 ms, but had difficulties at the original playback of a 100-ms burst period. For ITD conditions, the Pc did not exceed 55 %, independent of the burst period. Only for an ITD of 2400 μs at a 200-ms burst period, 60 % were reached, but this tendency of improvement has to be treated with caution.

Results for the RMR pitch task are presented on the left in Fig. 19. The effect of burst period, i.e. playback rate, is apparent. It was not possible for CI24 to complete the task at the original playback rate, but results are comparable to CI116 when conducted at half-tempo. The ITD task results (Fig. 19, right) lead to the conclusion that the RMR did not work for this listener, even at the largest ITDs under test. Fig. 20 displays the accumulated results of all pitch-ITD combination sets. It additionally features the 0-ITD data taken from the RMR pitch task for comparison. The positive effect of additional ITD cues on the segregation of streams

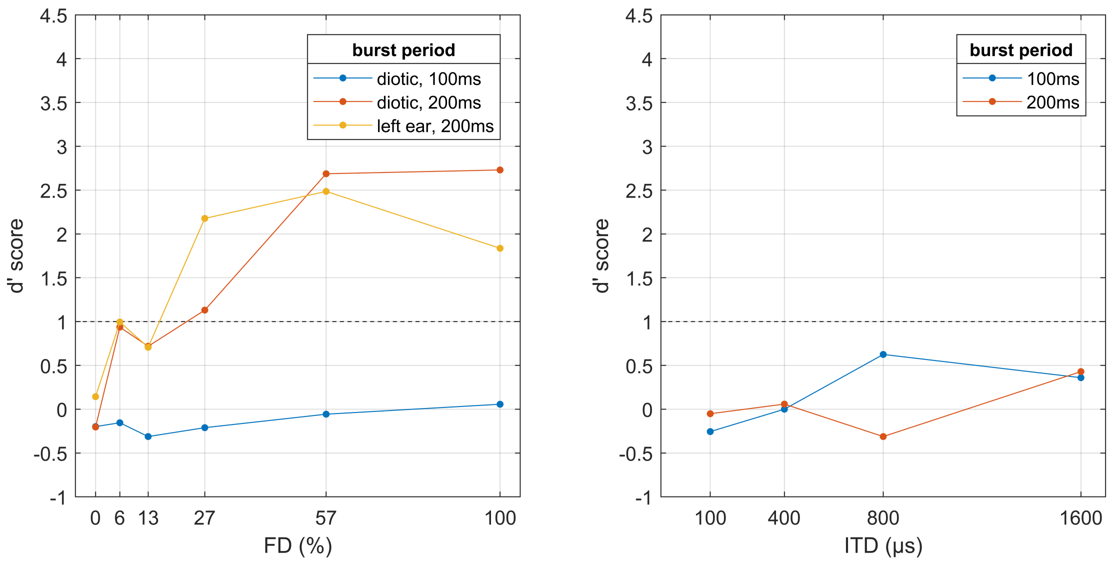


Figure 19: CI24: RMR d' scores. Pitch (left) and ITD (right) condition results for varying burst period.

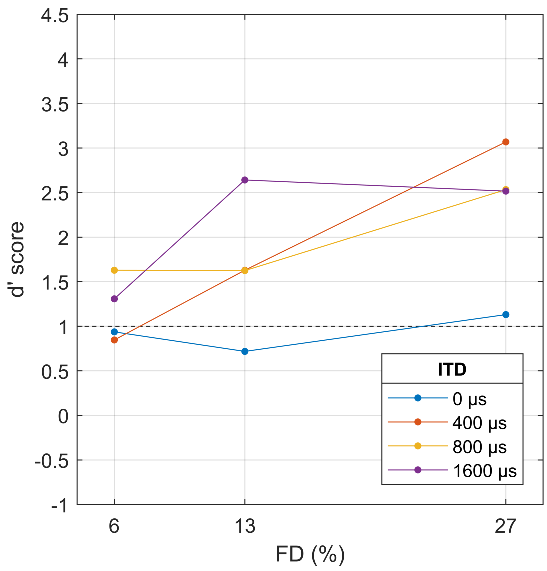


Figure 20: CI24: RMR d' scores. Pitch+ITD condition results for burst period 200 ms.

can be observed for all but one condition. Largest improvement of d' can be found for FDs of 13 and 27 %. Only nine combinations of FDs and ITDs could be tested by CI24 because of time constraints. Still, adding ITDs to 57 and 100 % FDs could have been promising for this listener as pitch task results did not reach ceiling scores.

2.4 Discussion and Conclusions

The pilot experiments were successful in terms of proving software functionality, confirming the feasibility of the RMR test design and even indicating synergistic tendencies for a combination of temporal cues. The listeners were able to segregate streams based on rate pitch cues, but could not do so using only ITDs. However, introducing ITDs to the smallest FDs substantially increased d' scores, even for the smallest ITD of 400 μs.

The loudness balancing procedure turned out to be rather time-consuming. To keep it as short as possible, in the future, balancing should only be conducted for the lowest and highest pitch rates with the amplitudes for the remaining rates linearly interpolated between the corresponding amplitudes and the CL of the nominal rate.

Furthermore, the results led to the conclusion that only LR conditions should be tested and only short stimuli should be used in the discrimination task, with an onset-to-onset time of 200 ms. CI116, although having a musical background and showing no trouble with the RMR task at the original playback rate, was concerned that playback could be too fast for other listeners. As the feedback of the pilot concerning improvements was highly regarded, conditions with slower rates were presented to CI24 and led to results in favor of a burst period of 200 ms instead of 100 ms also for all future participants.

Regarding the choice of ITD and FDs, differences in the individual listener abilities should be considered in the RMR tests, while maintaining a consistent overall testing schedule. To this end, a decision tree may help to plan evaluation of the formal training scores selecting testing sets for either average or high sensitivity to FDs and ITDs. It will be explained in detail in Sec. 3.1.

So far, RMR segregation was tested in a sequence of three blocks. To ensure that the benefit of the pitch-ITD combination was not only because of training effects from the prior sets, all conditions should be presented in a random order in one testing block (including scheduled breaks).

3 Main Experiment

The results and conclusions drawn from the pilot experiment were considered in the setup of the main experiment which was completed by bilateral CI listeners. The majority of them were regular participants of CI studies at the ARI or research facilities of their CI manufacturer, and were invited personally or through a call for participation via social media. In the following sections, the methods of the main experiment are described. Necessary modifications in conditions and procedures and the applied analysis of the obtained data are explained and supported by state-of-the-art practice described in psychoacoustic research literature. Subsequently, an extensive statistical analysis is conducted and results are presented in a graphical and numerical way.

3.1 Methods

3.1.1 Apparatus, Task and Stimuli

The same apparatus, task and stimuli were used as in the pilot experiments (see Sec. 2.1). In the main test, only short, low rate stimuli were tested in all blocks.

3.1.2 Participants

Seven bilaterally implanted listeners participated in the main experiment (6 female, 1 male). All of them were post-lingually deafened, CI114 and CI117 had a high level hearing loss since birth and used hearing aids before implantation. Each of the listeners was scheduled for two to three days in a row and received financial compensation for their participation. Tab. 5 lists individual subject information, including individual and average monaural (m) and binaural (b) electric listening experience and the single electrode pair that is stimulated during the experiment.

	Etiology of hearing loss	Age at testing	Age at deafness onset	CI experience L/R	Implant type[4] L	Implant type[4] R	Electrode Nr. L	Electrode Nr. R
CI1	meningitis	37 yr	14 yr	23/23 yr	C40+	Synchrony 2	7	6
CI12	progressive[5]	56 yr	29 yr	21/22 yr	C40+	C40+	6	6
CI17	idiopathic	75 yr	adult	16/5 yr	Synchrony	Synchrony	7	7
CI60	meningitis	74 yr	58 yr	16/16 yr	Pulsar	Pulsar	9	9
CI114	progressive	56 yr	39 yr	17/11 yr	Pulsar	Sonata	6	6
CI117	progressive	40 yr	adult	18/3yr	Synchrony	C40+	6	6
CI119	progressive	65 yr	adult	2/1 yr	Synchrony 2	Synchrony 2	6	6
avg		58 yr		m:16 yr, b:11 yr				

Table 5: Participant information including the average monaural (m) and binaural (b) electric listening experience.

3.1.3 Fitting

For the listeners having participated in previous experiments at the ARI, the utilized electrode pair was selected based on information from previous studies. For fist-time participants or for those who had never done any task concerned with ITDs, the first step was the selection of a sufficiently ITD-sensitive electrodes. This was the case for CI1, CI60 and CI117. Electrodes in the middle of the array were preferred to maintain comparability over all subjects of the present study to related prior and follow-up studies. Four combinations of the left and right electrodes 6 and 7, i.e., 6/6, 6/7, 7/6, and 7/7 were activated and a shortened version of the ITD discrimination task was conducted to see which pair enabled the best sensitivity. If neither of the results showed reliable ITD perception, a different pair in the near neighborhood should be tested until performance was promising. For these electrodes, THRs, MCLs, and CLs were determined and image centering was performed prior to the test at the nominal rate (118 pps, 8.5 ms) used throughout the ITD discrimination task. 50 repetitions per ITD condition (100, 400, 800, and 1600 μs) were presented, i.e., half the set of the main ITD discrimination task. Once the best electrode option was found, CLs were also checked for the highest and lowest rates and the loudness balancing and image centering continued as for the other participants.

[4]Manufactured by Med-EL GmbH in Innsbruck, Austria
[5]due to progressive vestibular aqueduct syndrome

Pitch$_{avg}$		Pitch$_{high}$		ITD$_{avg}$	ITD$_{high}$
pair (ms)	FD (%)	pair (ms)	FD (%)	ITD (μs)	ITD (μs)
8.5/8.5	0	8.5/8.5	0	100	100
9/8	13	9/8.5	6	500	200
9.5/7.5	27	9/8	13	1000	400
11/7	57	9.5/7.5	27	2000	800
12/6	100	11/7	57		
12/5	140	12/6	100		

Table 6: Pitch and ITD condition sets for average and high sensitivity.

3.1.4 Conditions

The FD (6, 13, 27, 57, and 100 %) and ITD (100, 400, 800, and 1600 μs) conditions for the pitch and ITD discrimination tasks were identical to the selection of the pilot experiment. All trials were presented at random order and breaks were included depending on the listener's individual endurance and focus. For CI1, CI60, and CI117 half of the repetitions of ITDs were already tested during the search for ITD sensitive electrode pairs. Thus, they completed the remaining part of the task and results were combined for the data analysis.

In the RMR task, FD and ITD conditions should be chosen according to the individual abilities a listener is exhibiting in the training phase for the task. To that end, sets of conditions were prepared, that either suit an average or high sensitivity to the presented cues. Table 6 lists the four options for the separate RMR pitch and RMR ITD task. For average ITD sensitivity, ITDs of 100, 500, 1000, and even 2000 μs were introduced. The latter did not fit the range where phase ambiguity could be fully avoided for the smallest pitch period used. However, the ITD that is applied here during the RMR task is present in the on- and offset of a burst, which supports the hearing in the evaluation, strongly reducing the ambiguity issue for the used stimuli. Considering that the participants have little to no experience with ITD cues, presenting exaggerated time differences could help them get more familiar with the new sensation itself, which could also be beneficial at smaller ITDs. Possible combinations forming the pitch-ITD combination sets 1-4 are displayed in Tab. 7, where all three ITD levels were added to each of the FDs.

Fig. 21 displays the decision tree that was installed and consulted in the selection

Set 1: Pitch$_{avg}$ / ITD$_{avg}$			**Set 2**: Pitch$_{high}$ / ITD$_{avg}$		
pair (ms)	FD (%)	ITD (μs)	pair (ms)	FD (%)	ITD (μs)
9/8	13	500/1000/2000	9/8.5	6	500/1000/2000
9.5/7.5	27		9/8	13	
11/7	57		9.5/7.5	27	
Set 3: Pitch$_{avg}$ / ITD$_{high}$			**Set 4**: Pitch$_{high}$ / ITD$_{high}$		
pair (ms)	FD (%)	ITD (μs)	pair (ms)	FD (%)	ITD (μs)
9/8	13	200/400/800	9/8.5	6	200/400/800
9.5/7.5	27		9/8	13	
11/7	57		9.5/7.5	27	

Table 7: Pitch-ITD combination sets 1-4. Selected ITDs were applied to all FD steps.

of the testing sets for each listener. The Pc score achieved in the newly introduced pitch-ITD condition (140 % FD, 2000 μs ITD) in the formal training was evaluated, with 61% being the first decision boundary to further inspect the Pc scores of the separate pitch and ITD training conditions. The sets for the main test were chosen according to the applied decision rules.

3.1.5 Tasks

Loudness Balancing and Image Centering As suggested by the results from the pilot experiment, the loudness balancing was shortened to a minimum to spare time for RMR and discrimination tasks. Only the amplitudes of the lowest and highest pitch rates used throughout were balanced via the adaptive procedure (Chapter 2, Sec. 2.2.1) and amplitudes for the remaining rates were obtained by linear interpolation. The subsequent process of informal centering of the auditory image was the same as in the pilot experiment.

Pitch and ITD Discrimination The tasks of pitch and ITD discrimination were identical to those from the pilot and were conducted prior to RMR tasks. The majority of the experienced listeners was already familiar with the discrimination procedures and controller handling which resulted in a less demanding start to the experiment on the first day of testing.

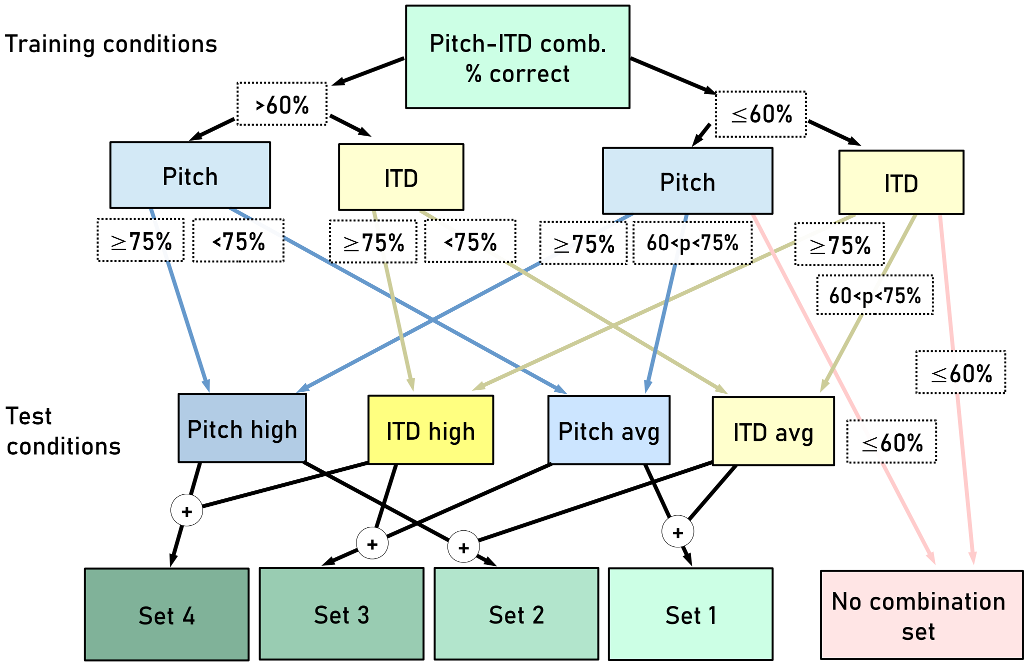


Figure 21: Decision tree: Pc (% correct) from the training conditions were used to select respective sets of test conditions for the three RMR tasks.

RMR Training Similar to the procedure in the pilot experiment, the main training started with the informal verbal instructions and the introduction to the stimuli. Particularly, focus was on showing all variations (orders) of one condition by repeating the assignments of higher and lower pitches and/or directions (ITDs) to the target and distractor streams multiple times. As the first bursts of a test sequence could appear from either the left or the right side and have either the higher or lower rate pitch, it was of great importance that all variations were known and could be correctly assigned to the individual rhythms by the listener.

After the introduction, the formal training without distractor stream was completed. All participants achieved fairly high scores (85 to 100 % correct identification), confirming their understanding of the basic rhythms. The last part of the training included the distractor bursts, and a pitch-ITD condition was complementing the existing conditions from the pilot setting. Table 8 shows the accumulated training results. Following the rules of the decision tree (Fig. 21), the same test sets were selected for all seven listeners: $\text{Pitch}_{\text{high}}$, ITD_{avg}, and the respective combination set 2 (see Tab. 6 and 7).

Training with distractor									
			% correct						
ITD (μs)	FD (%)	# trials	CI1	CI12	CI17	CI60	CI114	CI117	CI119
0	0	84	47.62	52.38	50	54.76	50	44.01	51.19
0	140	84	98.81	96.43	79.76	77.38	96.43	92.86	86.9
2000	0	84	52.38	61.9	41.67	48.81	54.76	60.71	54.76
2000	140	88	100	100	93.18	96.59	98.86	97.73	97.73
2500	0	84	-	-	-	-	57.14	48.81	-

Table 8: Results of the formal training with distractor (all listeners).

CI1	CI12	CI17	CI60	CI114	CI117	CI119
Pitch	Pitch	ITD	Pitch+ITD	Pitch	Pitch+ITD	ITD
ITD	ITD	Pitch	ITD	Pitch+ITD	Pitch	Pitch+ITD
Pitch+ITD	Pitch+ITD	Pitch+ITD	Pitch	ITD	ITD	Pitch

Table 9: Testing block order assigned to each individual listener.

RMR Test A complete randomized presentation of all conditions was incorporated in the main experiment to avoid the influence of the sequential order of the three (pitch, ITD, and pitch-ITD combination) testing blocks. However, randomization lead to utter confusion of the first listener (CI60) as they could not prepare themself for the quick changes in upcoming cues from one trial to the next. It could be assumed that this confusion might disappear with more extensive training and the listener might be able to adapt more quickly. Unfortunately, this could not be examined in the course of this study because of the participants' limited availability. As a compromise to avoid listening fatigue and cognitive overload in the main experiments, all RMR tasks were eventually presented in the three subsequent blocks, but block order was randomized between subjects. CI60 was re-invited to complete all RMR tasks for a second time to obtain new data. Table 9 shows the randomly assigned block order.

3.1.6 Data Analysis

The statistical analysis of the data consisted of significance tests (repeated-measures analysis of variance, **rmANOVAs**) and estimation of effect sizes η_G^2 (Bakeman 2005).

Calculations were performed in `Jamovi`[6] (Lenth 2020, R Core Team 2021, Singmann 2018, The jamovi project 2021) and MATLAB® in the versions 2017b or 2021b using the respective built-in statistical functions and additional toolboxes. Note that the repeated measures consider the individual performance of the listener within our small-N group design (Gouvea 2017, Holcombe 2021, Smith and Little 2018).The analysis approach stayed in line with procedures utilized in ongoing research related to the present study at the Acoustics Research Institute.

The significance level for the rmANOVA was determined at $\alpha = 5\,\%$. Oberfeld and Franke (2013) proposed the standard use of the Huynh-Feldt correction (Huynh and Feldt 1976) for the p and F values for small-N designs inherent to the field of CI research in psychoacoustics, as Huynh-Feldt corrected results are less prone to underestimate occurring effects compared to other corrections. Therefore, Huynh-Feldt correction is applied to all of the results provided in the tables of Sec. 3.2.

The size of an observed effect can be reported in the form of the generalized η_G^2 taking values from 0 to 1 (Olejnik and Algina 2003). It can be divided in ranges for small ($0.02 \leq \eta_G^2 < 0.13$), medium ($0.13 \leq \eta_G^2 < 0.26$) and large ($\eta_G^2 \geq 0.26$) effects and allows the interpretation of observations across within-subjects and between-subjects designs (Bakeman 2005). Due to η_G^2 representing an estimation, additional 90 %-confidence intervals of the effect size are retrieved by using the *ncpci* function of the `Measures of Effect Sizes Toolbox` (Hentschke and Stüttgen 2011).

[6]Statistical analysis spreadsheet application, based on R

3.2 Results

This section describes the resulting balanced and centered amplitudes from the pretests, followed by ANOVA tables and graphical display of d' scores for discrimination versus RMR. For the pitch-ITD combination, measurements results were compared to a prediction data set computed from the results of the pitch and ITD RMR performance to examine the hypothesized synergy effect.

3.2.1 Balanced and Centered Amplitudes

Current amplitudes of the ten pitch periods are listed for each listener in their respective stimulation range in Tab. 10. The amplitudes at 8.5 ms period (in *italics*) were the CLs obtained in the fitting, while **bold** amplitudes of the upper and lower bound were retrieved from the adaptive balancing.

3.2.2 Pitch: Discrimination versus RMR performance

The rmANOVA tables of this section show main effects of FD and the task type as well as their interaction. Significance levels and degrees of freedom are Huynh-Feldt corrected and 90 %-confidence intervals for the η_G^2 are listed in square brackets. Bold font is used for significant effects and considerable η_G^2 scores above 0.02. Calculations were performed on the basis of raw and normalized d' data, due to the two task types having different d' ceilings. Normalized d' were obtained by dividing d' by the $d'_{\max}$ for the respective task type, as they were described in Secs. 2.2.3, 2.2.4 and 2.2.5. Both metrics are considered important, because they represent different perspectives of the results.

Raw dataset For the raw data both effects and the interaction are highly significant. Effect sizes are large for the change of FD, almost medium for the task type but also medium for their interaction. Fig. 22 shows the graph for discrimination and RMR d' scores over the six FDs, each dot presenting a listeners' individual datapoint. The estimated marginal means (circles), standard errors (bars) and the bounds of the 95 % confidence intervals per FD are attached in Tab. 12 and were obtained via built-in functions of Jamovi.

Normalized dataset With data normalization, the effect of FD remained significant with a large effect size that is comparable to the raw data. However, the effect of the task and the interaction were not significant any more and showed a much smaller effect. Graphs are displayed in Fig. 23.

Balanced and Centered Amplitudes (cu)											
		Pitch period (ms)									
ID (range)		5	6	7	7.5	8	8.5	9	9.5	11	12
CI1 (R3)	Bal. L	**36**	38	39	40	41	*42*	42	41	40	**39**
	Bal. R	**45**	49	52	54	56	*58*	58	58	58	**58**
	Centr. L	37	38	39	40	41	42	41	41	39	41
	Centr. R	45	53	57	60	61	58	63	66	65	60
CI12 (R3)	Bal. L	**74**	77	80	81	83	*84*	87	91	101	**108**
	Bal. R	**58**	64	69	72	75	*78*	79	79	82	**83**
	Centr. L	71	77	79	80	75	84	87	89	96	102
	Centr. R	62	71	73	72	75	78	79	81	85	83
CI17 (R3)	Bal. L	**50**	49	47	46	46	*43*	47	48	54	**57**
	Bal. R	**57**	55	54	53	52	*57*	53	54	59	**62**
	Centr. L	51	51	48	46	47	45	47	49	55	57
	Centr. R	53	52	52	54	52	51	53	52	56	62
CI60 (R3)	Bal. L	**60**	63	66	68	69	*71*	71	71	71	**71**
	Bal. R	**66**	70	74	76	78	*80*	80	81	81	**82**
	Centr. L	68	62	66	68	69	71	71	71	71	72
	Centr. R	75	66	74	76	75	80	80	79	81	79
CI114 (R2)	Bal. L	**82**	88	95	98	101	*104*	104	103	103	**102**
	Bal. R	**78**	81	84	85	87	*88*	88	87	86	**85**
	Centr. L	85	88	95	98	101	104	104	103	103	102
	Centr. R	77	81	83	85	85	88	86	87	86	83
CI117 (R3)	Bal. L	**57**	58	60	61	61	*62*	62	62	63	**63**
	Bal. R	**39**	41	43	44	45	*46*	46	46	45	**45**
	Centr. L	57	58	60	61	61	62	62	62	64	63
	Centr. R	40	41	43	47	47	46	46	47	46	45
CI119 (R3)	Bal. L	**56**	58	61	62	63	*64*	65	65	68	**69**
	Bal. R	**66**	68	70	71	72	*73*	74	74	77	**78**
	Centr. L	59	62	64	66	66	64	64	67	71	69
	Centr. R	66	65	67	67	70	73	75	70	73	78

Table 10: Balanced and centered amplitudes of all listeners.

rmANOVA: Pitch, raw data					
Effect	F	df_{effect}	df_{error}	p_{HF}	η^2_G [5%, 95%]
FD (%)	48.62	3.52	28.12	**< 0.001**	**0.654** [0.544, 0.710]
Task	9.36	1.00	8.00	**0.016**	**0.128** [0.027, 0.256]
FD (%) × Task	5.66	2.48	19.84	**0.008**	**0.136** [0.031, 0.231]

Table 11: Discrimination versus RMR performance for FD (raw data).

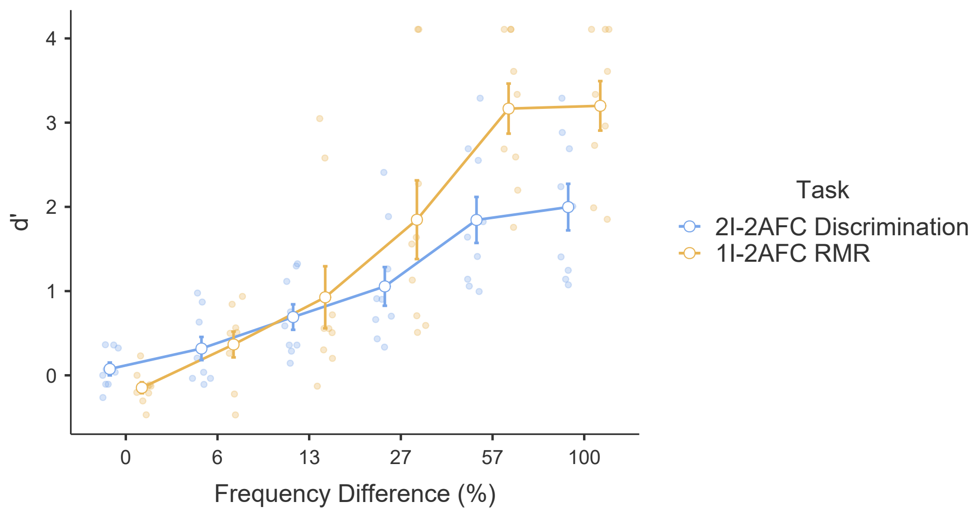


Figure 22: d' scores for FDs in discrimination and RMR tasks (raw data). Data (dots), estimated means (circles) and standard error for the mean (bars).

				95% Confidence Interval	
Task	FD (%)	Mean	SE	Lower	Upper
Discrimination	0	0.0754	0.0758	-0.0994	0.25024
	6	0.3181	0.1376	9.20e-4	0.63530
	13	0.6918	0.1512	0.3431	1.04049
	27	1.0557	0.2296	0.5263	1.58503
	57	1.8448	0.2731	1.2151	2.47444
	100	1.9974	0.2762	1.3605	2.63437
RMR	0	-0.1473	0.0648	-0.2966	0.00205
	6	0.3661	0.1531	0.0131	0.71914
	13	0.9266	0.3682	0.0776	1.77564
	27	1.8474	0.4671	0.7704	2.92449
	57	3.1667	0.2973	2.4811	3.85233
	100	3.1997	0.2939	2.5221	3.87732

Table 12: Estimated Marginal Means: FD versus task (raw data).

rmANOVA: Pitch, normalized data					
Effect	F	$\mathrm{df}_{\mathrm{effect}}$	$\mathrm{df}_{\mathrm{error}}$	p_{HF}	η^2_G [5%, 95%]
FD (%)	49.63	3.31	26.52	< **0.001**	**0.656** [0.544, 0.713]
Task	2.51	1.00	8.00	0.152	**0.038** [0 0.138]
FD (%) × Task	2.46	2.68	21.45	0.096	**0.061** [0, 0.130]

Table 13: Discrimination versus RMR performance for FD (normalized data).

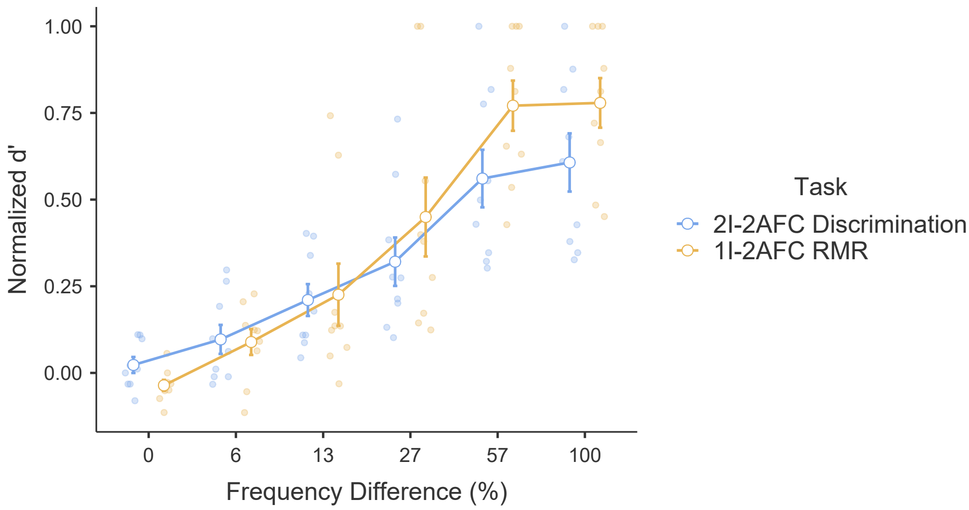


Figure 23: d' scores for FDs in discrimination and RMR tasks (normalized data). Data (dots), estimated means (circles) and standard error for the mean (bars).

				95% Confidence Interval	
Task	FD (%)	Mean	SE	Lower	Upper
Discrimination	0	0.0229	0.0230	-0.03020	0.0761
	6	0.0967	0.0418	2.80e-4	0.1931
	13	0.2103	0.0460	0.10428	0.3163
	27	0.3209	0.0698	0.15997	0.4818
	57	0.5607	0.0830	0.36934	0.7521
	100	0.6071	0.0840	0.41353	0.8007
RMR	0	-0.0359	0.0158	-0.07221	4.98e-4
	6	0.0891	0.0373	0.00320	0.1751
	13	0.2256	0.0896	0.01890	0.4323
	27	0.4498	0.1137	0.18756	0.7120
	57	0.7710	0.0724	0.60404	0.9379
	100	0.7790	0.0715	0.61402	0.9440

Table 14: Estimated Marginal Means: FD versus task (normalized data).

3.2.3 ITD: Discrimination versus RMR performance

Similar to the previous Sec. 3.2.2, rmANOVA tables and graphs display results for both unprocessed and normalized data.

Raw dataset ITDs had a significant effect which was represented by an η_G^2 of 0.248, classifying as a medium, almost large effect. The effect of the task and the interaction did not show significance in the p_{HF} value, but a very small effect was detected for the task choice. It is noted that the 95 %-confidence intervals in the last columns of Tab. 16 got below 0 for all ITD conditions of both tasks except for ITDs of 800/1000 μs and 1600/2000 μs in the discrimination task.

Normalized dataset Normalized data showed a significant effect of the ITD and at almost unchanged medium effect size. Further, the η_G^2 for effect of task observed a small effect, but not significant regarding p value.

rmANOVA: ITD, raw data					
Effect	F	df_{effect}	df_{error}	p_{HF}	η^2_G [5%, 95%]
ITD (μs)	8.120	1.91	13.40	**0.005**	**0.248** [0.072, 0.386]
Task	1.699	1.00	7.00	0.234	**0.044** [0, 0.184]
ITD (μs) × Task	0.212	1.76	12.35	0.786	0.006 [0, 0.048]

Table 15: rmANOVA results: discrimination versus RMR performance for ITD (raw data).

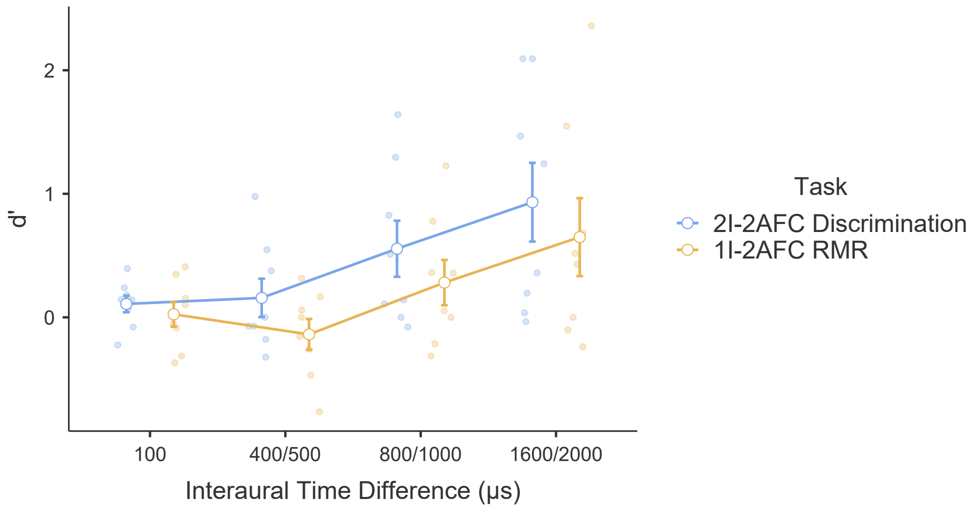


Figure 24: d' scores for ITDs in discrimination and RMR tasks (raw data). Data (dots), estimated means (circles) and standard error for the mean (bars).

				95% Confidence Interval	
Task	ITD (μs)	Mean	SE	Lower	Upper
Discrimination	100	0.1085	0.0673	-0.0506	0.268
	400/500	0.1575	0.1550	-0.2089	0.524
	800/1000	0.5555	0.2269	0.0191	1.092
	1600/2000	0.9317	0.3186	0.1784	1.685
RMR	100	0.0246	0.1000	-0.2118	0.261
	400/500	-0.1379	0.1248	-0.4331	0.157
	800/1000	0.2812	0.1833	-0.1522	0.715
	1600/2000	0.6496	0.3155	-0.0964	1.396

Table 16: Estimated Marginal Means: ITD versus task (raw data).

rmANOVA: ITD, normalized data					
Effect	F	df_{effect}	df_{error}	p_{HF}	η^2_G [5%, 95%]
ITD (μs)	8.842	1.91	13.35	**0.004**	**0.245** [0.078, 0.377]
Task	2.452	1.00	7.00	0.161	**0.071** [0, 0.234]
ITD (μs) × Task	0.534	1.78	12.46	0.579	0.014 [0, 0.073]

Table 17: rmANOVA results: discrimination versus RMR performance for ITD (normalized data).

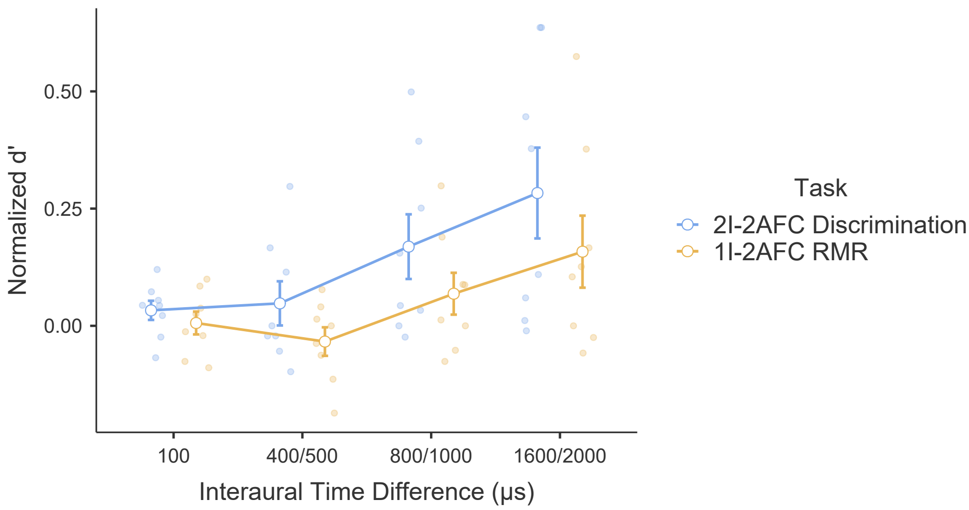


Figure 25: d' scores for ITDs in discrimination and RMR tasks (normalized data). Individual data (dots), estimated means (circles) and standard error for the mean (bars).

				95% Confidence Interval	
Task	ITD (μs)	Mean	SE	Lower	Upper
Discrimination	100	0.03298	0.0205	-0.01538	0.0813
	400/500	0.04787	0.0471	-0.06351	0.1593
	800/1000	0.16884	0.0690	0.00580	0.3319
	1600/2000	0.28321	0.0968	0.05423	0.5122
RMR	100	0.00599	0.0243	-0.05157	0.0635
	400/500	-0.03358	0.0304	-0.10544	0.0383
	800/1000	0.06846	0.0446	-0.03706	0.1740
	1600/2000	0.15814	0.0768	-0.02347	0.3398

Table 18: Estimated Marginal Means: ITD versus task (normalized data).

3.2.4 Pitch-ITD combination: RMR performance

The analysis in this section aims to determine if the combination of the two segregation cues lead to substantially better performance in the RMR task and therefore supports the hypothesis of an underlying synergy effect. The measurement results were compared to predictions created by adding the measurement results for pitch and ITD data, considering the varying sign of the d'. These calculations are based on Laneau et al. (2004) and McKay et al. (2000), assuming independent and optimally used cues. A normalization of the d' scores as in Sec. 3.2.2 and 3.2.3 was not necessary, because the compared data sets were of the same task type (RMR) and differed only slightly in the number of condition repetitions (96 versus 100). This is considered in the creation of the prediction data d'_{Pred} by limiting d' scores to the maximum d' for 96 repetitions, 4.0737:

$$X = [\mathrm{sign}(d'_{\mathrm{pitch}}) \cdot (d'_{\mathrm{pitch}})^2] + [\mathrm{sign}(d'_{\mathrm{ITD}}) \cdot (d'_{\mathrm{ITD}})^2] \tag{3.2.1}$$

$$d'_{\mathrm{pred}} = \min(\mathrm{sign}(X) \cdot \sqrt{|X|}, 4.0737) \tag{3.2.2}$$

To prove the synergy effect, measurement results should exceed the prediction. The respective statistics are listed in the rmANOVA Table 19 including all subject data under the effect of *Data Source*. This main effect was indeed significant with $p_{HF} =$ 0.016 and η^2_G could be interpreted as an effect of medium size, indicating synergy for the tested combinations of conditions. The effect of FD variation was highly significant and yielded a large η^2_G. Effect of ITD reaches significance as well, but the effect is rated as small. While the interaction of all three effects showed a marginal p_{HF} of 0.044, a negligible effect size of only 0.006 was estimated. Although it is not reaching significance, the interaction of *Data Source* $\times$ *FD* could be labeled as a small effect regarding η^2_G.

The 95 %-confidence intervals for each condition are listed in Tab. 20. With the exception of smallest ITDs applied to the smallest FD, all conditions have a lower bound above a d' of 0 which is indicating the chance level. Fig. 26 displays d' scores of the measurement and prediction data for the three ITD values per FD. Note that ITDs of CI24 (400/800/1600 μs) are included in the displayed groups and listed in Tab. 20 but are not mentioned in the labels of Fig. 26 due to better readability.

A three-way-interaction (Data Source $\times$ FD $\times$ ITD) was observed in the analysis that was on the edge of significance and did not show any effect in the sense of

rmANOVA: Pitch-ITD combination, raw data					
Effect	F	df_{effect}	df_{effect}	p_{HF}	η^2_G $[5\%, 95\%]$
Data Source	9.96	1.00	7.00	**0.016**	**0.141** [0.032, 0.273]
FD (%)	18.56	1.48	10.35	< **0.001**	**0.319** [0.154, 0.446]
ITD (μs)	10.73	1.43	10.04	**0.005**	**0.038** [0.012, 0.073]
Data Source × FD (%)	2.73	1.77	12.39	0.109	**0.025** [0, 0.067]
Data Source × ITD (μs)	1.49	2.00	14.00	0.258	0.002 [0, 0.007]
FD (%) × ITD (μs)	2.97	2.82	19.72	0.059	0.014 [0.001, 0.030]
Data Source × FD (%) × ITD (μs)	2.83	4.00	28.00	**0.044**	0.006 [0, 0.011]

Table 19: rmANOVA results: RMR performance for pitch-ITD combination (raw data).

					95% Confidence Interval	
FD (%)	Data Source	ITD (μs)	Mean	SE	Lower	Upper
6	Measurement	400/500	0.625	0.288	-0.05650	1.306
		800/1000	0.981	0.275	0.33195	1.631
		1600/2000	1.322	0.363	0.46374	2.179
	Prediction	400/500	0.226	0.213	-0.27813	0.730
		800/1000	0.548	0.219	0.03085	1.064
		1600/2000	0.902	0.298	0.19787	1.606
13	Measurement	400/500	1.781	0.372	0.90240	2.659
		800/1000	2.512	0.386	1.59950	3.424
		1600/2000	2.800	0.393	1.87125	3.728
	Prediction	400/500	0.978	0.410	0.00818	1.947
		800/1000	1.183	0.390	0.26147	2.104
		1600/2000	1.309	0.475	0.18620	2.431
27	Measurement	400/500	2.927	0.412	1.95361	3.900
		800/1000	2.682	0.366	1.81664	3.547
		1600/2000	3.053	0.418	2.06395	4.041
	Prediction	400/500	1.979	0.495	0.80888	3.149
		800/1000	2.048	0.490	0.88975	3.207
		1600/2000	2.087	0.490	0.92832	3.246

Table 20: Estimated Marginal Means: ITD versus data source versus FD.

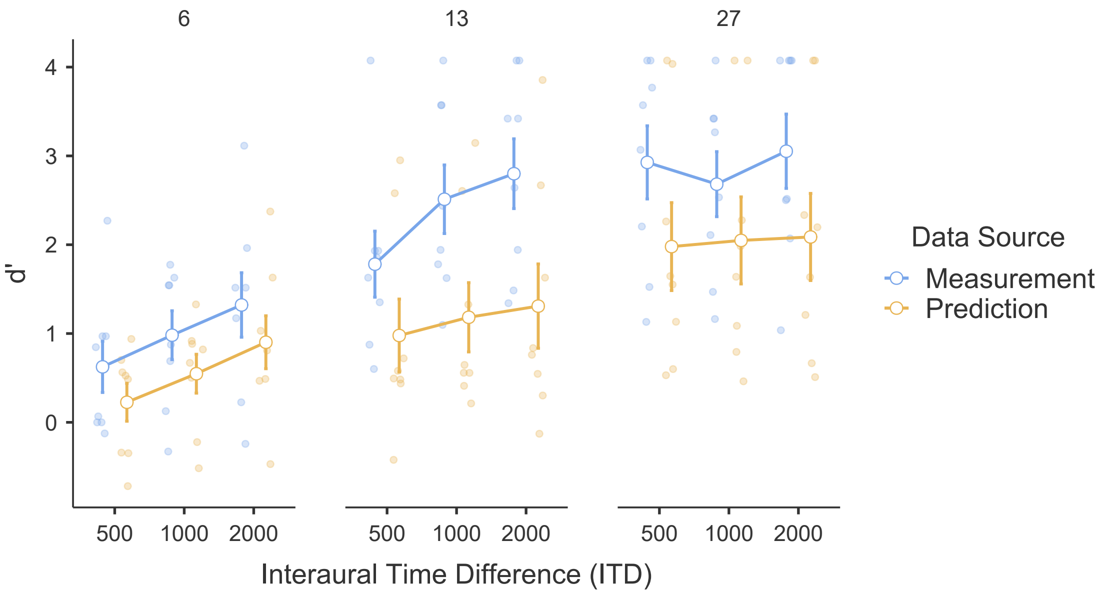
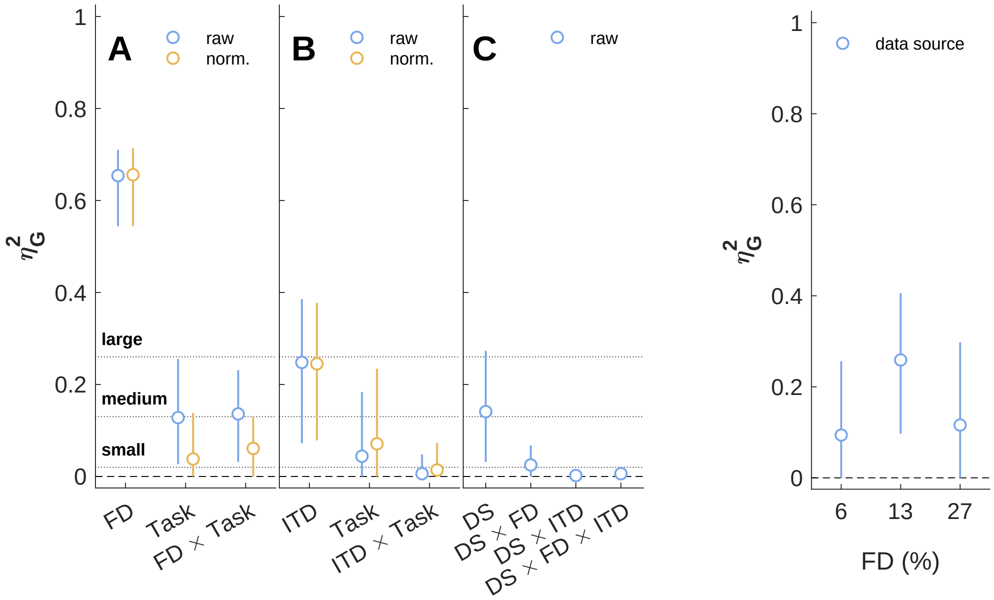


Figure 26: d' scores for measurement and prediction data for pitch-ITD combinations in the RMR task. Individual data (dots), estimated means (circles) and standard error for the mean (bars).

the η_G^2. As a follow-up, simple-effects rmANOVAs were conducted, i.e., analysis was repeated separately for each FD level and observed effects were therefore conditional on these levels. At this stage the numerical distinction between a small, medium, and large effect for η_G^2 should be avoided, because the obtained η_G^2 can not be directly compared to the η_G^2 of the preceding analyses that consider all factors, because of differences in their estimation. Therefore, η_G^2 should be only interpreted only on a relative level. Tab. 21 shows simple-effects rmANOVA results for *Data Source*, *ITD* and the interaction at the 6 % FD. Only the effect of ITD reached significance. In the case of the 13 % FD in Tab. 22, the rmANOVA showed significance for both effects and the interaction, with the latter presenting a much smaller η_G^2 relative to *Data Source* and *ITD* individually. For the largest FD of 27 % no significance could be observed for *Data Source* and the interaction, but for ITD (Tab. 23). Despite this outcome, estimated η_G^2 for ITD and the interaction are very close to zero while the *Data Source* effect is more than 20 times larger. Fig. 27a provides a visual representation of the estimated effect sizes including 90 %-confidence intervals for the analyzed data of all three RMR parts. Section A shows η_G^2 for pitch (from Tab. 11 and 13), section B for ITD (Tab. 15 and 17) and section C for *Data Source* (from



(a) Effect sizes for the effects of FD and task, ITD and task, and data source (DS) including interactions.

(b) Conditional effect sizes of data source effect, indicating synergy.

Figure 27: Effect size η_G^2 including 90% confidence intervals.

Tab. 19). Taking the analysis down one more level, rmANOVA was repeated again conditional at each FD, only considering the general effect of *Data Source* for any ITD at the FD level (labeled as simple-main-effect rmANOVA). No significance in the p_{HF} sense was reported for *Data Source* at the 6 % FD level (Tab. 24). For 13 % FD, however, the simple-main-effect analysis confirms the significance ($p = 0.004$) of the *Data Source* effect (Tab. 25). While not being significant at 27 %, *Data Source* presented a substantial η_G^2 (Tab. 26). Fig. 27b shows the estimated η_G^2 of this analysis level including confidence intervals.

Simple-effects rmANOVA: 6% FD					
Effect	F	$\mathrm{df}_{\mathrm{effect}}$	$\mathrm{df}_{\mathrm{error}}$	p_{HF}	η_G^2 [5%, 95%]
Data Source	3.6840	1.00	7.00	0.096	0.073 [0, 0.209]
ITD (μs)	5.6136	1.21	8.44	**0.039**	0.125 [0.013, 0.268]
Data Source × ITD (μs)	0.0104	2.00	14.00	0.990	0.000 [-]

Table 21: Follow-up simple-effects analysis, conditional at 6 % FD.

Simple-effects rmANOVA: 13% FD					
Effect	F	df_{effect}	df_{error}	p_{HF}	η^2_G $[5\%, 95\%]$
Data Source	17.17	1.00	7.00	**0.004**	0.241 [0.090, 0.383]
ITD (μs)	7.69	2.00	14.00	**0.006**	0.065 [0.019, 0.118]
Data Source × ITD (μs)	4.26	2.00	14.00	**0.036**	0.018 [0.002, 0.040]

Table 22: Follow-up simple-effects analysis, conditional at 13 % FD.

Simple-effects rmANOVA: 27% FD					
Effect	F	df_{effect}	df_{error}	p_{HF}	η^2_G $[5\%, 95\%]$
Data Source	3.84	1.00	7.00	0.091	0.114 [0, 0.294]
ITD (μs)	4.14	2.00	14.00	**0.039**	0.005 [0.001, 0.012]
Data Source × ITD (μs)	2.94	2.00	14.00	0.086	0.004 [0, 0.010]

Table 23: Follow-up simple-effects analysis, conditional at 27 % FD.

Simple-main-effects rmANOVA: 6% FD					
Effect	F	df_{effect}	df_{error}	p_{HF}	η^2_G $[5\%, 95\%]$
Data Source	3.68	1.00	7.00	0.096	0.094 [0, 0.256]

Table 24: Follow-up simple-main-effects analysis, conditional at 6 % FD.

Simple-main-effects rmANOVA: 13% FD					
Effect	F	df_{effect}	df_{error}	p_{HF}	η^2_G $[5\%, 95\%]$
Data Source	17.2	1.00	7.00	**0.004**	0.259 [0.097, 0.406]

Table 25: Follow-up simple-main-effects analysis, conditional at 13 % FD.

Simple-main-effects rmANOVA: 27% FD					
Effect	F	df_{effect}	df_{error}	p_{HF}	η^2_G $[5\%, 95\%]$
Data Source	3.84	1.00	7.00	0.091	0.116 [0, 0.298]

Table 26: Follow-up simple-main-effects analysis, conditional at 27 % FD.

4 Discussion

The detailed analysis of the accumulated data provides a solid basis to address the research questions formulated in Sec. 1.5. The feasibility of the RMR paradigm for CI studies, the actual effects of temporal cues on voluntary segregation, and the relation between sensitivity and segregation thresholds are assessed in the following.

CI listeners were indeed able to perform the adapted RMR task based on rate pitch cues for most of the presented FDs, with a strong effect of the rate in the tested range from 83 to 168 pps. Overall performance was above chance, ranging from as little as 13 % FD to a plateau between FDs of 57 % and 100 %. The FDs needed for segregation were even smaller than those of Paredes-Gallardo et al. (2018), who observed segregation for FDs only greater that 50 % in an overall testing range from 80 to 300 pps.

Surprisingly, the results of the pitch discrimination revealed a more shallow slope with increasing FD, although starting slightly above the RMR performance for the smallest FDs. The trend for better RMR performance at larger FDs remained even after normalization of d' scores. The statistics revealed an effect of the task type, leading to the assumption that RMR, and therefore voluntary segregation can be achieved more easily at larger FDs compared to the listeners' limits of the observed discrimination ability. This, however, is opposed to suggestions of Paredes-Gallardo et al. (2018) and Hong and Turner (2009) who state that larger differences are needed for segregation than for discrimination of sounds. One explanation for the differences in performance could be the short signals in the discrimination task (50 ms per burst) as opposed to the accumulated duration of all the bursts in an RMR rhythm sequence, where cues can be integrated over time. Hence, some listeners might have needed a longer duration in the discrimination task to reach a performance equivalent to that from the segregation task. While constraints in testing time did not leave room for such explorations, longer burst durations are worth considering in follow-up experiments. Yet, in this thesis, the independence of the sensitivity to FDs in the discrimination task and the performance in the segregation task cannot be ruled out.

Results for ITDs are not as distinct as they are for rate pitch. ITD perception was a great challenge for most of the listeners, because they have never been confronted with such cues in everyday life. Despite overall poor to moderate d' scores, a beneficial effect of large ITD has been observed in both raw and normalized data. The overall performance in RMR was above chance for ITDs of 800 μs and more, while in the discrimination task, the performance was above chance for all ITDs. Although the effect of the task was not significant in the p_{HF} sense, the effect size η_G^2 still indicated a small effect because the performance on discrimination was better for all of the presented ITDs. Assuming that the hypothesis about the different processing pathways for spatial cues in NH (Middlebrooks and Onsan 2012) also applies in the electric hearing, this might provide a partial explanation of the differences between discrimination and segregation results. Furthermore, the lack of training time could have been a reason for the weak performance in the demanding experiment. Participants shared the opinion that training with ITDs over the course of a few weeks to months could be beneficial to the outcome of the RMR tests, as most of them reported a definite presence of a marginal directional percept that was hard to grasp and process, but tended to manifest with repeated stimulation.

The applied burst period of 200 ms, i.e., the presented rate of bursts, was at the border for the occurrence of streaming in NH (Middlebrooks and Onsan 2012). However, the CI listeners expressed their need for the 200-ms presentation rate to be able to follow the rhythmic patterns. For barely perceived cues like ITDs, a longer burst period (and therefore a longer overall duration of the rhythm sequence) might be even more important to achieve better scores in the RMR task. This observation is in line with Paredes-Gallardo et al. (2018), who reported the possibility of a slower segregation build-up in electric hearing.

The combination of FD and ITD cues in the RMR task led to exciting results: The analysis of measurement and prediction data revealed a synergy effect of pitch and ITDs. In detail, an increase in both pitch and ITD led to a stronger segregation. While FDs showed a clear effect, ITDs held only a small share in the total effect size. By individual evaluation of each FD step, greatest synergy was found for the FD of 13 %, which is close to the musical interval of a whole tone. For the three tested FDs, the amount of presented ITD played a significant role, and at an FD of 6 %, the effect of ITD weighted even more than the effect of the data source, i.e., the indicator for synergy.

Despite having a small impact on segregation on their own, ITD cues seem to have served as a synergistic addition for the well-perceived temporal pitch cues. A possible explanation for this was hypothesized by Darwin (1997), who suggested that ITD may enable strong sequential grouping in complex sounds like speech, when the secondary information of location is present in streams that have already been segregated based on more salient cues like pitch. Consequently, a listener may be able to actively attend to the established location, which in turn results in even better segregation. Moreover, the binaural "sluggishness" of the auditory system (Middlebrooks and Onsan 2012) when following sudden binaural changes might have also played a role regarding the impact of the ITD cues of the target and distractor streams in the conducted RMR experiments.

While the applied RMR methodology was based on Middlebrooks and Onsan (2012), it was also influenced by the works of Paredes-Gallardo et al. (2018) who, to the best of the author's knowledge, were the only ones to unite an objective task of voluntary stream segregation with direct stimulation via a research interface. By incorporating a temporally irregular distractor stream in a delay detection task, they were hoping for a more likely occurrence of segregation. However, this introduced unwanted gap cues between certain target and distractor bursts which might have helped the listeners solving the task. Even as this cue emerged as a minor influence in the evaluation, the expected benefit of the distractor irregularity could not be observed. Therefore, following the concept of deterministic target and distractor streams in the adapted RMR tasks used in this thesis held no disadvantage, but helped avoiding the additional gap cues. Still, the RMR methodology cannot entirely rule out that the listeners might have relied on mechanisms other than segregation to complete the tasks. For instance, they might have attended to the changes in the continuous stream of bursts, alternating either in frequency or the direction the singular bursts were perceived from. To counterbalance said eventuality, sequence training and verbal instructions included the presentation of the basic rhythms on their own. Moreover, the participants were asked for their feedback and reported hearing two streams in the pitch tasks, however, this was not always the case in the ITD tasks. In fact, small ITDs seemed to be hard to grasp, but the majority of the listeners were still able to noticed some kind of lateral shift, despite having trouble verbalizing their exact sensations.

5 Conclusions and Outlook

In this master's thesis, CI listeners' stream segregation ability based on temporal cues was explored. For this purpose, the rhythmic masking release (RMR) paradigm was adapted for bilateral CI listeners, then tested and validated in listening experiments. The participants' segregation ability was compared to their performance in discrimination tasks. The evaluation involved low-rate frequency differences varying between 83 and 168 pps and ITDs ranging from physiologically plausible 100 μs to exaggerated 2000 μs. These cues were applied to the target and distractor streams (in the stream segregation task) or two consecutive trial intervals (in the discrimination task).

The general results revealed that CI listeners were capable of sequential stream segregation by exploiting the presented cues. While the effect of rate pitch was apparent, large ITDs were required to show any effect. The simultaneous presentation of both cues resulted in a better segregation performance compared to that achieved by the presentation of individual cues, exceeding the hypothetical performance obtained from the addition of the individual pitch and ITD task performances. This indicates that having access to both temporal cues yields a synergy effect in the task of stream segregation. When translating this significant benefit to CI users, the accessibility to salient rate pitch and ITD information should be particularly considered in the development of new stimulation strategies.

Even without extensive training, CI listeners were able to achieve high sensitivity in the rate-pitch experiments. In the ITD experiments, though, the duration of the training phase seems to play an important role. Additional training sets over a long time period might help the listeners to better process the interaural temporal information. This may facilitate ITD-based segregation for ITDs being close to the natural limit of around 800 μs.

In future experiments, the methodology applied in this thesis can be further improved. For example, by introducing a formal subjective task complementing the objective measure of the RMR, it may be ensured that segregation is used as the key mechanism for completing the task. Further, the informal image centering can be refined and even extended to a formal task. Moreover, it would be interesting to examine the RMR performance of bilateral listeners for high-rate stimuli as these – due to time and logistic constraints – were not tested in this thesis in favor of low-rate stimuli. In such high-rate conditions, ITDs would not only be a strong cue present at the onset, but would also be conveyed in the amplitude modulations of the signal. Another interesting research condition is the coupling of ITD and *place* pitch cues or even the combination of ITD, rate pitch, and place pitch. These conditions were already considered in the development of the testing software, paving the road for future experiments.

References

Bakeman, R. (2005), 'Recommended effect size statistics for repeated measures designs', *Behavior Research Methods* **37**(3), 379–384.

Beauvois, M. W. and Meddis, R. (1997), 'Time decay of auditory stream biasing', *Perception & Psychophysics* **59**(1), 81–86.

Bernstein, J. G., Jensen, K. K., Stakhovskaya, O. A., Noble, J. H., Hoa, M., Kim, H. J., Shih, R., Kolberg, E., Cleary, M. and Goupell, M. J. (2021), 'Interaural place-of-stimulation mismatch estimates using CT scans and binaural perception, but not pitch, are consistent in cochlear-implant users', *Journal of Neuroscience* **41**(49), 10161–10178.

Blauert, J. (1997), *Spatial Hearing: The Psychophysics of Human Sound Localization*, MIT press.

Bregman, A. S. (1978), 'Auditory streaming is cumulative', *Journal of Experimental Psychology: Human Perception and Performance* **4**(3), 380.

Bregman, A. S. and Campbell, J. (1971), 'Primary auditory stream segregation and perception of order in rapid sequences of tones', *Journal of Experimental Psychology* **89**(2), 244.

Chatterjee, M., Sarampalis, A. and Oba, S. I. (2006), 'Auditory stream segregation with cochlear implants: A preliminary report', *Hearing Research* **222**(1-2), 100–107.

Cooper, H. R. and Roberts, B. (2009), 'Auditory stream segregation in cochlear implant listeners: Measures based on temporal discrimination and interleaved melody recognition', *The Journal of the Acoustical Society of America* **126**(4), 1975–1987.

Darwin, C. J. (1997), 'Auditory grouping', *Trends in Cognitive Sciences* **1**(9), 327–333.

David, M., Lavandier, M. and Grimault, N. (2015), 'Sequential streaming, binaural cues and lateralization', *The Journal of the Acoustical Society of America* **138**(6), 3500–3512.

Duran, S. I., Collins, L. M. and Throckmorton, C. S. (2012), 'Stream segregation on a single electrode as a function of pulse rate in cochlear implant listeners', *The Journal of the Acoustical Society of America* **132**(6), 3849–3855.

Fearn, R., Carter, P. and Wolfe, J. (1999), 'The perception of pitch by users of cochlear implants: Possible significance for rate and place theories of pitch', *Acoustics Australia* **27**(2), 41–43.

Frohmann, L. (2022), 'SIPI loudness: Experimental setup and pilot study for loudness perception in cochlea-implant listeners with short-interpulse-interval (SIPI) and enhanced (ENH) pulse stimulation'.

Füllgrabe, C. and Moore, B. C. J. (2012), 'Objective and subjective measures of pure-tone stream segregation based on interaural time differences', *Hearing Research* **291**(1-2), 24–33.

Füllgrabe, C. and Moore, B. C. J. (2014), 'Effects of age and hearing loss on stream segregation based on interaural time differences', *The Journal of the Acoustical Society of America* **136**(2), EL185–EL191.

Gouvea, J. (2017), 'Insights from small-N studies', *CBE - Life Sciences Education* **16**(3), fe4.

Green, D. M., Swets, J. A. et al. (1966), *Signal Detection Theory and Psychophysics*, Vol. 1, Wiley New York.

Hartmann, W. M. and Macaulay, E. J. (2014), 'Anatomical limits on interaural time differences: An ecological perspective', *Frontiers in Neuroscience* **8**, 34.

Hentschke, H. and Stüttgen, M. C. (2011), 'Computation of measures of effect size for neuroscience data sets', *European Journal of Neuroscience* **34**(12), 1887–1894.

Holcombe, A. O. (2021), 'When average is over: small N but many trials.', *Journal of Cognition* **4**(1).

Hong, R. S. and Turner, C. W. (2006), 'Pure-tone auditory stream segregation and speech perception in noise in cochlear implant recipients', *The Journal of the Acoustical Society of America* **120**(1), 360–374.

Hong, R. S. and Turner, C. W. (2009), 'Sequential stream segregation using temporal periodicity cues in cochlear implant recipients', *The Journal of the Acoustical Society of America* **126**(1), 291–299.

Hu, H. and Dietz, M. (2015), 'Comparison of interaural electrode pairing methods for bilateral cochlear implants', *Trends in Hearing* **19**, 2331216515617143.

Huynh, H. and Feldt, L. S. (1976), 'Estimation of the box correction for degrees of freedom from sample data in randomized block and split-plot designs', *Journal of Educational Statistics* **1**(1), 69–82.

Ihlefeld, A., Carlyon, R. P., Kan, A., Churchill, T. H. and Litovsky, R. Y. (2015), 'Limitations on monaural and binaural temporal processing in bilateral cochlear implant listeners', *The Journal of the Acoustical Society of America* **16**(5), 641–652.

Jesteadt, W. (1980), 'An adaptive procedure for subjective judgments', *Perception & Psychophysics* .

Kan, A. and Litovsky, R. Y. (2015), 'Binaural hearing with electrical stimulation', *Hearing Research* **322**, 127–137.

Klein, S. A. (2001), 'Measuring, estimating, and understanding the psychometric function: A commentary', *Perception & Psychophysics* **63**(8), 1421–1455.

Kong, Y.-Y., Deeks, J. M., Axon, P. R. and Carlyon, R. P. (2009), 'Limits of temporal pitch in cochlear implants', *The Journal of the Acoustical Society of America* **125**(3), 1649–1657.

Laback, B., Egger, K. and Majdak, P. (2015), 'Perception and coding of interaural time differences with bilateral cochlear implants', *Hearing Research* **322**, 138–150.

Laneau, J., Wouters, J. and Moonen, M. (2004), 'Relative contributions of temporal and place pitch cues to fundamental frequency discrimination in cochlear implantees', *The Journal of the Acoustical Society of America* **116**(6), 3606–3619.

Lenth, R. (2020), 'emmeans: Estimated marginal means, aka least-squares means.', [R package].
URL: *Retrieved from https://cran.r-project.org/package=emmeans*

Lindenbeck, M. J. (2018), Temporal pitch in electric hearing with short-interpulse-interval stimulation, Master's thesis, University of Technology Graz, University of Music and Performing Arts Graz.

Litovsky, R. Y., Colburn, H. S., Yost, W. A. and Guzman, S. J. (1999), 'The precedence effect', *The Journal of the Acoustical Society of America* **106**(4), 1633–1654.

Majdak, P., Laback, B. and Baumgartner, W.-D. (2006), 'Effects of interaural time differences in fine structure and envelope on lateral discrimination in electric hearing', *The Journal of the Acoustical Society of America* **120**(4), 2190–2201.

McKay, C. M., McDermott, H. J. and Carlyon, R. P. (2000), 'Place and temporal cues in pitch perception: Are they truly independent?', *Acoustics Research Letters Online* **1**(1), 25–30.

Middlebrooks, J. C. and Onsan, Z. A. (2012), 'Stream segregation with high spatial acuity', *The Journal of the Acoustical Society of America* **132**(6), 3896–3911.

Mihocic, M., Majdak, P. and Laback, B. (2012), 'ExpSuite: Software-Framework zur Durchführung psychoakustischer Experimente', *presented at the 62. Jahrestagung der Österreichischen Physikalischen Gesellschaft (ÖPG), Graz* .

Moore, B. C. J. and Carlyon, R. P. (2005), Perception of pitch by people with cochlear hearing loss and by cochlear implant users, *in* 'Pitch', Springer, pp. 234–277.

Moore, B. C. J. and Gockel, H. (2002), 'Factors influencing sequential stream segregation', *Acta Acustica United with Acustica* **88**(3), 320–333.

Oberfeld, D. and Franke, T. (2013), 'Evaluating the robustness of repeated measures analyses: The case of small sample sizes and nonnormal data', *Behavior Research Methods* **45**(3), 792–812.

Olejnik, S. and Algina, J. (2003), 'Generalized eta and omega squared statistics: Measures of effect size for some common research designs.', *Psychological Methods* **8**(4), 434.

Paredes-Gallardo, A., Madsen, S. M. K., Dau, T. and Marozeau, J. (2018), 'The role of temporal cues in voluntary stream segregation for cochlear implant users', *Trends in Hearing* **22**.

Plack, C. J. (2018), *The Sense of Hearing*, Routledge.

R Core Team (2021), 'R: A language and environment for statistical computing.', Version 4.0 [Computer software].
URL: *https://cran.r-project.org. (R packages retrieved from MRAN snapshot 2021-04-01)*

Robles, L. and Ruggero, M. A. (2001), 'Mechanics of the mammalian cochlea', *Physiological Reviews* **81**(3), 1305–1352.

Sach, A. J. and Bailey, P. J. (2004), 'Some characteristics of auditory spatial attention revealed using rhythmic masking release', *Perception & Psychophysics* **66**(8), 1379–1387.

Seebeck, A. (1844), 'Über die Definition des Tones', *Annalen der Physik* **139**(11), 353–368.

Shannon, R. V. (1983), 'Multichannel electrical stimulation of the auditory nerve in man. I. Basic psychophysics', *Hearing Research* **11**(2), 157–189.

Singmann, H. (2018), 'afex: Analysis of factorial experiments.', [R package].
URL: *https://cran.r-project.org/package=afex.*

Smith, P. L. and Little, D. R. (2018), 'Small is beautiful: In defense of the small-N design', *Psychonomic Bulletin & Review* **25**(6), 2083–2101.

Stainsby, T. H., Füllgrabe, C., Flanagan, H. J., Waldman, S. K. and Moore, B. C. J. (2011), 'Sequential streaming due to manipulation of interaural time differences', *The Journal of the Acoustical Society of America* **130**(2), 904–914.

Tejani, V. D., Schvartz-Leyzac, K. C. and Chatterjee, M. (2017), 'Sequential stream segregation in normally-hearing and cochlear-implant listeners', *The Journal of the Acoustical Society of America* **141**(1), 50–64.

The jamovi project (2021), 'jamovi', Version 2.2 [Computer Software].
URL: *https://www.jamovi.org*

Townshend, B., Cotter, N., Van Compernolle, D. and White, R. L. (1987), 'Pitch perception by cochlear implant subjects', *The Journal of the Acoustical Society of America* **82**(1), 106–115.

Turgeon, M., Bregman, A. S. and Ahad, P. A. (2002), 'Rhythmic masking release: Contribution of cues for perceptual organization to the cross-spectral fusion of concurrent narrow-band noises', *The Journal of the Acoustical Society of America* **111**(4), 1819–1831.

Van Hoesel, R. J. M. (2004), 'Exploring the benefits of bilateral cochlear implants', *Audiology and Neurotology* **9**(4), 234–246.

Van Noorden, L. P. A. S. (1975), 'Temporal coherence and the perception of temporal position in tone sequences', *IPO Annual Progress Report* **10**, 4–18.

Verschooten, E., Shamma, S., Oxenham, A. J., Moore, B. C. J., Joris, P. X., Heinz, M. G. and Plack, C. J. (2019), 'The upper frequency limit for the use of phase locking to code temporal fine structure in humans: A compilation of viewpoints', *Hearing Research* **377**, 109–121.

Vliegen, J., Moore, B. C. J. and Oxenham, A. J. (1999), 'The role of spectral and periodicity cues in auditory stream segregation, measured using a temporal discrimination task', *The Journal of the Acoustical Society of America* **106**(2), 938–945.

Vliegen, J. and Oxenham, A. J. (1999), 'Sequential stream segregation in the absence of spectral cues', *The Journal of the Acoustical Society of America* **105**(1), 339–346.

Von Békésy, G. and Wever, E. G. (1960), *Experiments in Hearing*, Vol. 195, McGraw-Hill New York.

Von Helmholtz, H. (1954), 'Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik (F. Vieweg und Sohn, Braunschweig, Germany). Translated as: On the Sensations of Tone as a Physiological Basis for the Theory of Music, by A.J. Ellis from the 4th German edition, 1877, Leymans, London, 1885'.

Wilson, B. S. and Dorman, M. F. (2008), 'Cochlear implants: A remarkable past and a brilliant future', *Hearing Research* **242**(0), 3–21.

Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., Eddington, D. K. and Rabinowitz, W. M. (1991), 'Better speech recognition with cochlear implants', *Nature* **352**(6332), 236–238.

Wilson, B. S., Finley, C. C., Lawson, D. T. and Zerbi, M. (1997), 'Temporal representations with cochlear implants', *American Journal of Otology* **18**, 30–34.

List of Figures

List of Tables