Chapter 1

INTRODUCTION

1.1 PROBLEM STATEMENT

1.1.1 Context of the problem

Cochlear implant (CI) devices afford many profoundly deaf individuals worldwide partially restored hearing ability (Wilson & Dorman 2008). Although CI users achieve remarkable speech perception with contemporary multichannel CIs (Gifford et al. 2008), reports of generally unsatisfactory music perception (for example Gfeller et al. 1997; Leal et al. 2003; Looi et al. 2004; McDermott 2004; Gfeller et al. 2005; Looi et al. 2008) have prompted research efforts directed at improving CI users' music listening experience.

One of the main challenges facing scientists involved in CI research is that the input delivered to the central auditory system after peripheral electrical stimulation is deprived compared to that for normal hearing (NH). Owing to only 22 electrodes being available to handle much of the frequency range relevant to human auditory tasks, acoustic–electrical signal conversion results in severely limited spectral resolution, compounded by mismatched frequency-to-place mapping (Fu & Shannon 2002) and channel interaction (Hughes & Abbas 2006). However, CI-mediated perception of temporal cues, as measured during gap detection tasks, is comparable to that of NH listeners (Shannon 1989). In a functionally hierarchically organised perceptual system, such as the auditory system (Pandya 1995; Kaas et al. 1999; Wessinger et al. 2001; Zatorre et al. 2002; Griffiths & Warren 2002; Semple & Scott 2003; Stewart et al. 2008; Warren 2008; Woods & Alain

2009), the outcome of events at earlier levels serves as input to subsequent processing levels. It follows that impoverished input deriving from an early level will propagate through the entire processing system, possibly to the detriment of ultimate perceptual outcome. For successful CI-mediated hearing it is therefore necessary not only to encode stimulus attributes accurately during acoustic–electrical conversion, but also to know their perceptual outcomes.

This context is pertinent to understanding CI-mediated perception, and specifically music perception, in real-world listening conditions. Successful perception of specifically Western tonal music relies largely on accurate processing of spectral information for pitch perception, but with important contributions from temporal cues for the perception of rhythm and metre. Music perception has consequently been proposed to be facilitated by a modular, hierarchically organised processing system that initially consists of two independent but parallel streams (Peretz & Coltheart 2003), of which the outputs later combine to create a unified musical percept (Peretz 1990). In order to improve music perception experience for CI listeners, it is thus necessary to understand how impoverished information – both as isolated cues and in interaction – translate to perceptual outcome in real-world listening conditions (Middlebrooks et al. 2005; Wilson & Dorman 2008; Moore & Shannon 2009). Such understanding rests on three pillars: (i) understanding the underlying requirements for and mechanisms of music perception to allow relevant information to be extracted peripherally and presented for further auditory processing, (ii) linking perceptual experience of music-relevant stimuli in real-world listening conditions to the psychophysical ability afforded by the implant device and (iii) investigating and understanding CI-mediated music perception as the outcome of systems-based processing of an auditory input.

1.1.2 Research gap

Improved CI-mediated music perception ability requires that the underlying constraints hindering processing of music-relevant information need to be understood. This implies the need for systematic investigation of CI-mediated perception of simple music-relevant stimuli, as isolated elements as well as in combination. Several previous studies have addressed CI-mediated perception of music-relevant stimuli, but have focused on using either isolated cues (for example Gfeller & Lansing 1991; Gfeller & Lansing 1992; Pijl & Schwarz 1995b; Pijl 1997; Gfeller et al. 1997; Kong et al. 2004) or familiar melodies comprising complex tones and combined pitch, rhythm and timbre information (for example Gfeller et al. 2003; Looi et al. 2004; Gfeller et al. 2005; Vongpaisal et al. 2009). Although such studies provide insightful comment regarding CI users' limited music perception ability, results were not interpreted in the context of a complete sensory perceptual system as operates in real-world listening nor related to the psychophysical ability allowed by the implant technology. Since CI sound processors are generally developed to extract speech cues, progress towards improving specific processing deficits underlying poor music perception will remain elusive without understanding the constraints related to CI-mediated extraction of musical cues from auditory input, their presentation to the processing system and subsequent cognitive processing. By considering music as consisting of multiple components that are presented to the auditory periphery as a unified whole, an approach to understanding its perception by deconstructing the elements as well as combining them, may improve our understanding of how elementary perceptual ability afforded by the CI device may translate into real-world like melodic sequences.

Earlier studies regarding CI-mediated music perception, especially those aimed at investigating real-world music listening ability, have often used familiar melodies as stimuli during assessment of CI listeners music perception ability (see for example Gfeller et al. 1997; Leal et al. 2003; Kong et al. 2004; McDermott & Looi 2004; Gfeller et al. 2005; Looi et al. 2008). Listeners are usually asked to identify a melody from a list of well-known tunes following an adjustment to the sound processing algorithm or the listener's map settings. However, owing to only a set number of response choices being available a listener may have to draw on several cognitive processing strategies, including mnemonic analysis, to support the response decision. This is not an undesirable ability in real-world listening; however, when it forms the basis of a task to probe music listening success, the behavioural outcome may not be able to be linked unambiguously to perception of a specific component of the input. In this sense the task of music perception then is rather one of music recognition, meaning that music perception ability is assessed in relation to *familiarity with a music excerpt*, rather than the successful recognition of specific musical features. Again, results from such investigations contribute to assessing CI users' music listening behaviour, but do not facilitate a deeper understanding of the underlying processing that facilitates the observed behaviour.

Whether CI users experience a sequence of successive tones of specific frequency and duration and separated by fixed silences as melody-like has not been tested previously. Yet, by using familiar melodies as a probe to gauge CI listeners' melody perception ability, it is assumed that they do. It may thus be valuable to take one step back in the quest to understand CI-mediated music perception and first determine whether the output of signal processing at the periphery, which serves as input to cortical processing, contains sufficient auditory information to allow a listener to infer melodic quality of a tone sequence.

Implant users' desire for restored music listening ability may prompt successful music perception to become the next benchmark for CI success (Limb et al. 2010). To this end, two clinically relevant assessment tools, which both offer valuable entry points to the assessment of CI-mediated music perception, have recently been proposed (Cooper et al. 2008; Kang et al. 2009). However, given the multifaceted nature of music as auditory stimulus, objectively gauging the real-world perceptual effect of manipulations to processing algorithms, user settings or biophysical constraints will require an understanding not only of the perception of deconstructed musical features but also of their interaction within a complete perceptual system. This study thus aims to put forward a systematic investigation into CI-mediated music perception by investigating perception of specific cues, in isolation and combination, as they may contribute to music perception. Use of pure tone signals presented in sound-field listening conditions allows perceptual outcomes to be interpreted in consideration with signal processing strategies as well as systems-based processing of electrically delivered auditory information. The approach thus allows an intermediate step to assess perception of a subjective auditory quality in as objective a manner as possible in order to facilitate improved understanding of CI-mediated music perception (also see paragraph 1.5 and Figure 1.2).

1.2 BACKGROUND LITERATURE REVIEW

Figure 1.1 provides a schematic representation of the respective study fields that inform the understanding of music perception in electrical hearing. Broadly, these fields can be clustered into two streams, one relating to the understanding of music perception in NH and the other to the study of perceptual ability following electrical peripheral auditory stimulation. Each is informed by an understanding of the underlying neurobiology of hearing to in turn inform the understanding of CI-mediated music perception. It is important to note that the schema is not an exhaustive representation of the processing modules that contribute to a listener's forming a final, unified musical percept, but rather a conceptual model that guided the thought process during the current study. Although an important part of music perception, timbre perception is not included in the schema as it was not probed as one of the empirical investigations during the course of the study. The schema is anchored in the model put forward by (Peretz & Coltheart 2003), which provides a more comprehensive overview of processing modules that form part of the music perception pathway.

The next sections will provide a brief literature review to sketch the context within which the relevance of specific components of the schema can be understood. Emphasis is specifically on components included in this schema; the literature review does therefore not attempt to provide a comprehensive discussion of extant literature regarding music perception in either normal or electrically mediated hearing. Authoritative and expansive reviews regarding these aspects already exist (Deutsch 1998; Loizou 1999; Krumhansl 2000; Peretz & Zatorre 2003; McDermott 2004; Rubinstein 2004; Limb 2006a).

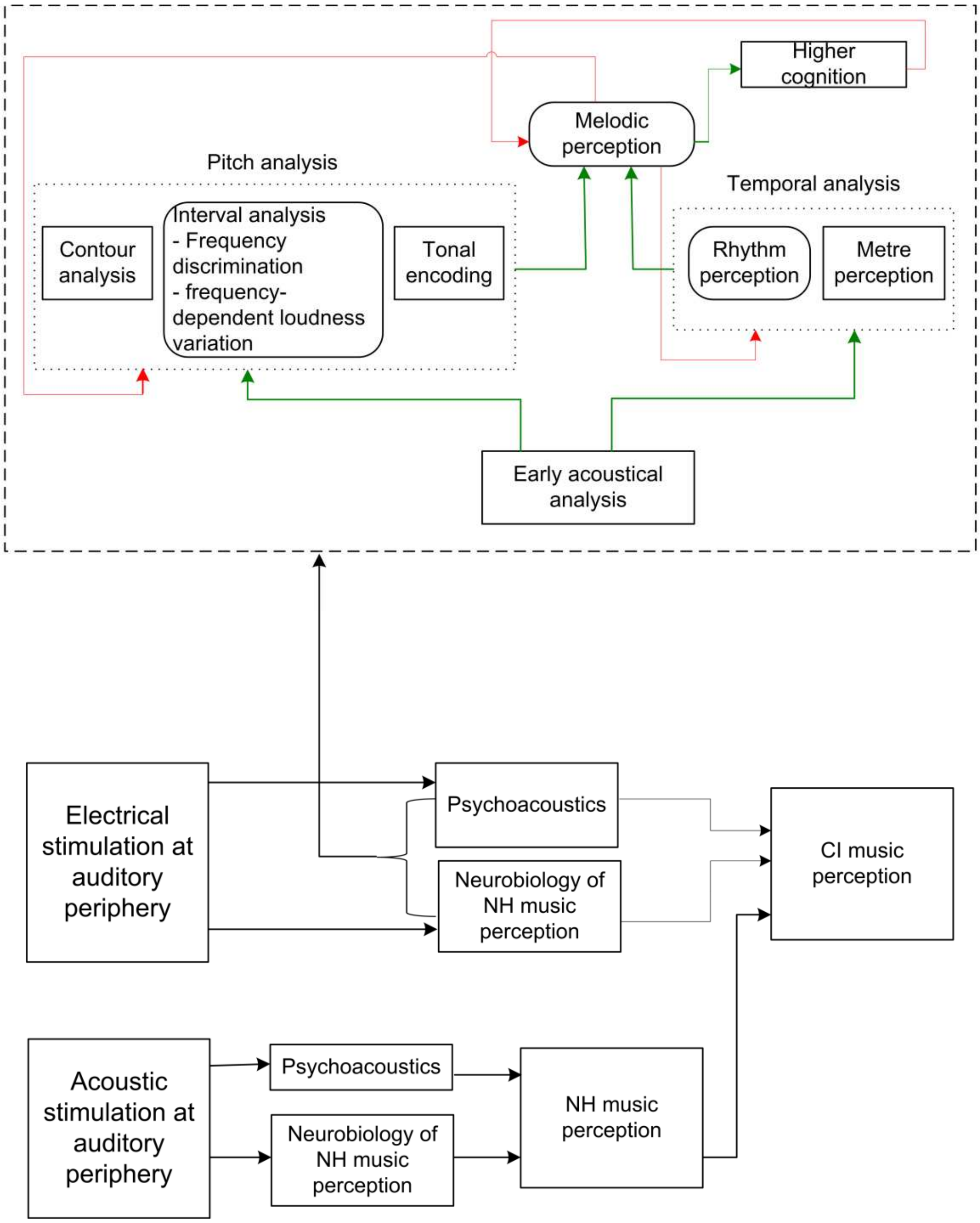

Figure 1.1: Schematic representation of the context of the study. The approach to improve understanding of CI music perception is informed by knowledge regarding psychoacoustics and neurobiology underlying music perception. The conceptual framework on which systematic investigation in this study is based is developed in the dashed frame. Processing modules that contribute to independent pitch and time-related analysis are enclosed in dotted frames. Resulting output feeds into a processing module at a higher hierarchical level, from where output is sent further again. Green arrows represent forward information flow; red arrows represent modulatory feedback. The rounded blocks indicate specific processing steps investigated during this study.

1.2.1 Music as auditory input

Owing to the substantial acoustic similarity between speech and music it is difficult to define why music is a unique auditory input. Both stimuli employ sequences of successive frequency changes that develop over time to communicate meaning according to a rule-based syntax (Limb 2006a), yet music is accompanied by an aesthetically pleasurable experience quite different from language. Perhaps the distinction resides in the perceptual meaning of music being conveyed through the relationship of individual tones to preceding and following ones within the context of the presented sequence (Peretz & Zatorre 2005). As such music is probably best defined as a collection of sound patterns built on the physical dimensions frequency and duration to convey perceptually meaningful pitch and time-related information organised according to the principles of pitch, rhythm and harmony (Krumhansl 2000; Gray et al. 2001; Limb 2006b).

Although pitch and rhythm are posited as the primary dimensions of music (Krumhansl 2000), successful music perception appears to rely foremost on analysis of frequency-related information (Zatorre 2001). Recent studies have shown that the inability to recognise music is the result of pitch perception deficits at various levels (Peretz 2002; Peretz & Hyde 2003; Foxton et al. 2004), which provides empirical support for the central position afforded to pitch in a theoretical framework of musical structure (Krumhansl 1990). According to this framework, different pitches, each associated with a specific frequency, are organised along musical scales. Within a scale each pitch is in turn associated with a tonal priority as dictated by the implied melodic scale (Krumhansl 1990), and so the relevance of a specific tonal pitch (and thus the associated frequency) within a tonal sequence depends on the relationship between frequencies rather than absolute frequencies per se (Krumhansl 1990; Trainor et al. 2002; Peretz & Zatorre 2005). The size of each melodic pitch interval is thus associated with a fixed frequency ratio (Hartmann 1998).

Similar to the tonal dimension, temporal cues in music (i.e. metre and rhythm) also find perceptual relevance only when duration of tones and neighbouring silences are heard in relation to those of the rest of the sequence (Peretz 1990; Peretz & Zatorre 2005).

However, unlike tonal information, temporal information does not convey the defining characteristics of music, but rather facilitates and supports music perception at an organisational level. This is in line with Jones's (1993) theory of joint accent structure, which posits that temporal accents coinciding with tonal accents increase perceptual saliency of melodic moments and so promote melody perception. The notion of temporal information fulfilling a supporting role during music perception has been underlined by studies which showed CI listeners' recognition of familiar melodies to improve in the presence of tonal as well as temporal cues, as opposed to recognition of melodies when only the pitch contour was preserved (Kong et al. 2004; Galvin et al. 2007).

1.2.2 Neurobiological foundations of music perception

Music perception is believed to be underlain by a modular cortical processing architecture, evident at various organisational levels (Peretz & Coltheart 2003; Koelsch & Siebel 2005). Firstly, processing of music and language information appears to be performed by largely independent processing networks despite some localised functional overlap between cortical areas. This view is based on collective data from studies with amusic subjects, who, following cortical lesions, presented with impaired music perception abilities while speech perception remained intact (Peretz et al. 1994; Nicholson et al. 2003).

The notion of a modular processing architecture underlying music processing is further supported by observations that cortical lesions may lead to impaired pitch-related perception without compromising the perception of temporal properties of music, and vice versa. As shown in Figure 1.1, the approach suggests that music processing modules are organised along two independent subsystems operating in parallel – one handling the pitch content of the acoustic input and the other dealing with the temporal content – that consist of several individual, hierarchically organised information processing steps (Peretz & Coltheart 2003; Koelsch & Siebel 2005; Warren 2008). The systems have been shown to be hierarchically organised with regard to both anatomy and function; elementary spectrotemporal properties appear to be processed in primary auditory cortex, followed by analyses of relationships between successive acoustic events, short-term retention of these patterns and other higher cognitive functions in a widely distributed network beyond primary auditory cortex (Zatorre et al. 1994; Platel et al. 1997; Patterson et al. 2002; Scott & Johnsrude 2003; Limb 2006b; Warren 2008).

The functional distinction between music- and language processing networks, as well as the distinction between pitch- and rhythm processing modules may be related to functional hemispheric specialisation. Observations of perceptual deficits with regard to tonal and temporal aspects of music in patients with unilateral brain damage (Peretz & Babaï 1992; Schuppert et al. 2000), together with results from neuroimaging studies (Zatorre et al. 1994; Liégeois-Chauvel et al. 1998; Schonwiesner et al. 2005), point to possible hemispheric specialisation for processing of tonal and temporal properties of complex sounds. On closer inspection, such lateralised specialisation appears to be linked to the differential contribution from temporal and spectral properties to speech and music and presents a possible framework according to which shared functionality between the two domains can be explained. Given that properties relevant to speech recognition are mostly contained in rapid temporal variations, while fine spectral structure conveys music-relevant information (Peretz & Hyde 2003; Zatorre 2003), a processing system that can handle both input types with similar precision would be required. Owing to the inverse relationship between frequency and time, improved spectral or temporal resolution will come at the expense of resolution in the complementary dimension if both were to be handled by a processing system wired only for one dimension. Zatorre and colleagues (2002) suggested that hemispheric specialisation developed to allow the auditory system to process both speech and music with the necessary precision. Such a hypothesis is supported by neuroimaging studies that show complementary responsiveness to fine temporal and spectral variations in the left and right auditory cortex respectively (Zatorre & Belin 2001). It thus appears that lateralisation is not directed at speech and music perception per se, but rather at functions dealing with temporal and spectral processing that, as argued earlier, form the underlying basis of the distinction between speech and music as auditory input.

1.2.3 Perceptual ability afforded by the CI device

In the case of sensorineural deafness, the cochlea, responsible for converting acoustic signals to a neurally encoded representation, is severely impaired (Loizou 1999). Auditory information is therefore not represented accurately and may be severely degraded, which in turn affects the fidelity with which the auditory nerve conveys information to the central auditory nervous system. Cochlear implants aim to circumvent this problem by decomposing a sound signal into its frequency components and subsequently stimulating the auditory nerve electrically.

The device consists of two units: the (speech) signal processor worn externally, and an implanted receiver–stimulator unit. The external components consist of a microphone and speech processor, usually worn behind the ear, and a transmitting coil. These are responsible for converting acoustic signals to electrical stimuli and subsequent transmission of the signals to the internal components. The internal unit is responsible for electrical stimulation of the auditory nerve and consists of a receiver coil, a decoder that generates pulsatile electrical stimuli (both implanted behind the ear), and an electrode array inserted in the scala tympani to stimulate auditory nerve fibres originating in the cochlea.

Place pitch has been shown to be an important pitch perception mechanism (Oxenham et al. 2004) and may be especially relevant for successful CI-mediated music listening (Townshend et al. 1987; Swanson et al. 2009), as purely temporal pitch is available only to about 300 Hz (Zeng 2002). Yet in CI-mediated hearing the resolution of the place pitch mechanism is governed by 22 discrete electrodes placed, in the case of the Nucleus device, at 0.75 mm intervals on the basilar membrane and reaching maximally only about two-thirds into the cochlea. Each electrode is associated with a frequency band, of which the width, in turn, is determined by the filter bank chosen for the specific signal processing strategy. For the Nucleus device the smallest filter width equals 125 Hz, and is applied only to the first eight or ten electrodes. Later electrodes (i.e. coding for higher frequencies) are associated with wider filters (and thus stimulatory frequency ranges). The place pitch mechanism in CI-mediated hearing thus allows only limited spectral resolution, which is probably reduced even further by factors such as filter overlap and unfocused current distribution (Hughes & Abbas 2006). The combined effect of acoustic–electrical conversion, reduced spectral resolution and mismatched frequency-to-place mapping (Fu & Shannon 2002) severely hinders CI users’ pitch perception ability.

Such coarse spectral resolution does not support CI-mediated music perception ability (Shannon 2005), especially when one considers that in almost 70% of all melodic intervals in Western tonal music, the pitch difference between two successive tones appears not to exceed two semitones (Vos & Troost 1989), translating to a frequency ratio of only 1:1.122. Since perception of the relative pitch changes between successive tones is so important to music perception, it follows that a perceptual system without sufficient pitch processing resolution will be unable to process an essential part of musical structure (Peretz & Hyde 2003). This is possibly one of the main reasons for poor music perception abilities of CI users.

Acoustic–electrical conversion, however, does yield sufficient information to support adequate speech perception. Speech appears to be fairly robust to spectral degradation (Shannon et al. 1995), probably since speech cues contained in spectral fine structure can be conveyed adequately through coarse spectral recoding. Further properties relevant to speech recognition are contained in rapid temporal variations. Taken together this implies that speech perception requires temporal cues to be delivered with high fidelity and less focus on fine spectral resolution (Shannon 2005).

Insights regarding hemispheric specialisation (as discussed earlier) and the associated processing benefits can advance the understanding of CI-mediated music perception. For example, based on the work of Mazziotta et al. (1982, as cited in Platel et al. 1997) and Phelps and Mazziota (1985, as cited in Platel et al. 1997), Platel et al. (1997) argue that difficult pitch perception tasks during music listening may induce analytical listening strategies subserved by left hemisphere activity. Considering then that speech perception (and per implication temporal processing) is subserved predominantly by left hemisphere structures, that several speech areas have been implicated in music perception (Koelsch et al. 2002; Levitin & Menon 2003) and that CI users receive regular activation of speech structures, it is possible that CI users' left hemispheric speech structures may be accustomed to receiving and dealing with impoverished information. This may in turn facilitate analytical processing to some extent during CI listeners' music perception when rhythmic cues (Fujita & Ito 1999; Kong et al. 2004) and lyrics (Fujita & Ito 1999; Gfeller et al. 2003) are available.

Recent imaging studies by Koelsch et al. (2004) and Limb et al. (2010) provide further support for the relevance of understanding processing constraints of the CI system in the context of the underlying neurobiology. Both studies showed that similar cortical structures are activated during CI-mediated music perception as for NH listeners, but that the pattern and intensity of activation differ. These findings underline that music perception is possible in electrical hearing, but that the quality of the input signal influences processing at cortical level, especially since the music processing system is posited to be subject to both bottom-up input and top-down feedback control (Purwins et al. 2008).

1.3 RESEARCH OBJECTIVES

1.3.1 Hypothesis

The hypothesis of this study is that investigation of CI-mediated music listening with simple yet musically relevant stimuli in sound-field conditions can provide useful insights that may allow perceptual outcome to be linked to the psychophysical abilities afforded by the CI device. Such insights may help to guide systematic efforts at improving signal processing strategies. The specific research questions put forward form only a limited part of a wider research effort regarding CI users’ music perception abilities and experiences.

Figure 1.1 is a schematic representation of the thought process guiding the study. It is important to note that research endeavours regarding music perception in normal and electric hearing have not developed in parallel owing to each discipline addressing research questions specific to behavioural outcome experienced by the distinct end users. However, as shown in Figure 1.1, investigation into CI-mediated music perception can draw on a similar approach, using results from psychoacoustic experiments as well as insights regarding the underlying neurobiological mechanisms, to inform understanding of CI music perception in the context realistic listening conditions confronting a CI user.

1.3.2 Research questions

The context described in the earlier literature review and developed schematically in Figure 1.1 serves to guide formulation of research questions which address elements of both pitch and temporal analysis during CI-mediated music perception. Recombination of the two dimensions after initial separation at an earlier level in the processing hierarchy follows during melody perception. Output from this stage feeds forward to higher-order cognition.

The investigation aims to put forward an approach to provide a deeper understanding of factors that contribute to CI listeners' ability to perceive music-relevant auditory information. Systematic investigation of information processing as occurring in sound-field conditions allows CI-mediated perception of musical information to be interpreted with consideration to a hierarchically organised perceptual system, and so may provide a link, in future studies, for inferring perceptual outcome from adjustments that apply to processing strategy design.

Research questions relate to elements shown in rounded blocks within the dashed frame in Figure 1.1.

1. Considering that successful melody perception in the Western tonal tradition requires the pitch changes of successive tonal events to be interpreted in relation to those of flanking events (Krumhansl 2000; Peretz & Hyde 2003), what is the typical frequency discrimination ability of CI users in sound field-listening conditions, specifically measured over an extensive range of base frequencies?[1] Can the perceptual outcome be explained with regard to signal processor design, and if so, to which design features? How would considering CI-mediated perceptual

---

[1] At the time when this research question was drafted and the investigation initiated (2005), such frequency discrimination data had not been available.

outcome as the result of systems-based processing contribute to the understanding of perception of music-relevant information in electric hearing?

2. An earlier study regarding amusic listeners' rhythm perception abilities showed significantly poorer perception of pitch-varying rhythmic patterns than experienced by NH listeners (Foxton et al. 2006). Given that pitch and temporal information are posited to be handled by independent processing streams (Peretz & Coltheart 2003), a pitch processing deficit was not expected to influence rhythm perception ability. However, the results of the study by Foxton et al. (2006) showed that adding pitch complexity to tone sequences adversely affected amusic listeners' rhythm perception abilities. Since CI listeners can be regarded as functionally amusic owing to pitch processing constraints, it raises the question of whether their rhythm perception abilities are influenced to the same extent as that of amusic listeners owing to pitch processing constraints. Therefore, how do CI listeners typically perform on rhythm perception tasks when confronted with simple tone patterns in sound-field conditions when pitch and rhythm cues covary?

3. Music as a perceptual input is difficult to define. Investigating its perception is complicated by not knowing whether the listener experiences a tone sequence as a melody. Even more difficult is to assess perceptual success with regard to musical input if the accuracy of transfer of components that contribute to forming a musical percept is unknown. To improve the ability of the CI device to convey music-relevant information it is necessary to determine whether the CI device conveys sufficient information to facilitate perception of the melody-like character of a tone sequence. How can the transfer of melodic character during electrically mediated hearing be probed objectively? To what extent can CI listeners judge musical characteristics as defined by musical syntactic congruency in short Western tonal melodies compared to NH listeners? How can such insights contribute to the understanding of CI-mediated music perception?

4. With reference to the relationship between pitch and intensity perception (Stevens 1935) and its relevance in electric hearing (e.g. Arnoldner et al. 2006), how much

frequency-dependent loudness variation do CI listeners typically experience when presented with musically relevant stimuli in sound-field conditions? How does this compare with that experienced by NH listeners?

1.3.3 Approach

The specific research questions put forward were addressed through perceptual experiments (see dashed frame in Figure 1.1). To this end, task-specific stimuli and new test procedures, specifically for sound-field conditions, were developed. Since the ultimate goal of cochlear implantation is to create an artificial perceptual system that mimics the normal auditory system, results obtained from CI users in the respective tasks were compared to those from NH listeners to allow meaningful interpretation of systems-based CI-mediated music perception. Specific methodological designs as applied to address the respective research questions are discussed in more detail in relevant later chapters.

1.4 RESEARCH CONTRIBUTION

According to the context outlined in section 1.3, this study focussed on the systematic investigation of aspects that may further current understanding of CI-mediated music perception in sound-field listening conditions. Specific research contributions are listed below.

1. Sound-field frequency discrimination thresholds for CI users were determined across an extensive range of base frequencies, with specific consideration to the position of the base frequency relative along the frequency response curves associated with the filter bank. Results showed that, in listening conditions that resemble those experienced in daily life, the CI technology as is currently available may provide pitch resolution that is better than what would be expected with place coding of pitch, which is considered to be governed by the output of the filter bank to listeners. Findings were interpreted with regard to signal processor filter

properties. The results were encouraging considering the importance of fine pitch resolution for successful perception of Western tonal music and may help to guide future signal processing strategies aimed specifically at improving CI-mediated music perception. Typical thresholds were in line with results by Gfeller et al. (2007). The work was published in *Hearing Research* (see Pretorius & Hanekom 2008).

2. Rhythm perception abilities were investigated using music-relevant tone sequences under conditions of varying pitch and rhythmic complexity. This allowed the specific effect of a varying pitch pattern on rhythm perception ability to be investigated. Compared to results for amusic listeners (Foxton et al. 2006) in a similar study, the observations point to a possible influence of the position at which an auditory deficit originates on the extent of constrained perceptual outcome.

   The experimental design also allowed investigation of CI users' ability to perceive rhythmic patterns in the context of a tone sequence. The findings may help to further contemporary understanding of CI users' ability to perceive and understand music in realistic listening conditions. The work is currently under review for publication in *Cochlear Implants International* (see Pretorius et al. 2010).

3. CI listeners' ability to perceive musical character as conveyed during electrically mediated hearing was investigated using tasks that probed their ability to judge syntactic congruency of simple, single-voice melodies. Using novel melodies allowed CI listeners' perception of real-life melodies to be investigated without the confounding constraints associated with using familiar musical excerpts. The work presents a first approach to developing test modules to gauge whether sufficient music-relevant information is conveyed to support successful melody perception. Upon further refinement such modules may add and complement existing tests that assess CI-mediated music perception success. An article about this work has been submitted to peer review (see Pretorius & Hanekom 2011).

4. Investigation into the frequency-dependent loudness variation CI listeners experience when listening through their own processors in quiet but realistic listening conditions appeared to be similar to that experienced by NH listeners. The finding suggests that clinically assigned processor settings adequately mimic loudness perception abilities. The findings may help to guide further investigation into the influence of unbalanced loudness cues during music listening, for comment on the relevance of rigorous loudness balancing prior to commencing with music perception tests in sound-field listening conditions.

1.5 OVERVIEW OF THE STUDY

As described earlier (paragraph 1.2) the understanding of music perception in electric hearing is broadly informed by knowledge regarding music perception in NH and investigation into perceptual ability following electrical stimulation of the auditory periphery. The present study focused on the latter as shown in Figure 1.1.

As can be seen from Figure 1.1 the study is only a partial investigation of aspects that contribute to successful music perception. They are bound together by their functioning within the greater processing system underlying CI-mediated music perception and specifically since all investigations were performed in sound-field listening conditions. The approach presents an intermediate step in a hierarchy of possible approaches to reach a better understanding of music perception facilitated by electric hearing.

Figure 1.2 shows that existing findings from basic psychoacoustic experiments, conducted both at the electrode level and in sound-field listening conditions, were used as starting point to inform the present study. Investigation into basic psychoacoustic abilities of CI users when confronted with simple cues in sound field conditions, whether in isolation or in combined form in melody-like tone sequences, presents an intermediate level in the approach hierarchy.

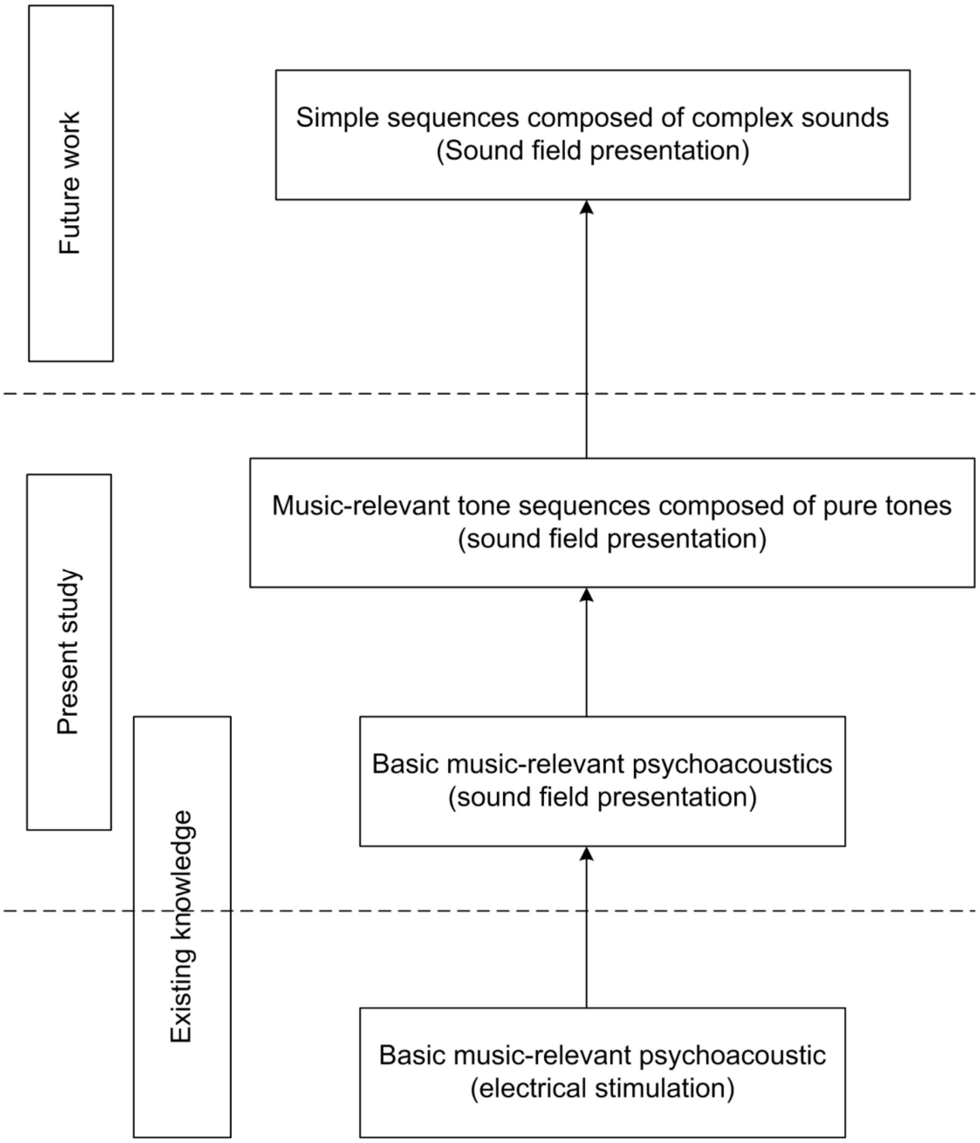

Figure 1.2: Hierarchical approach for studying CI-mediated music perception. The respective investigations reported in this study represent an intermediate step to link findings from electrode-level investigations to real-world perceptual outcomes.

In **Chapter 2** sound-field frequency discrimination was investigated and findings were interpreted with regard to speech processor design. Outcomes from investigation at this level can inform study at a subsequent level where basic cues are combined into simple, musically relevant stimuli presented in sound-field conditions. This was implemented during the investigation of CI users' rhythmic perception ability with regard to simple tones sequences where pitch and rhythm cues covaried (**Chapter 3**). The extent to which combined basic cues as provided by electric hearing facilitate melodic judgements was investigated in **Chapter 4**. The perceptual relationship between pitch and loudness and prompted a small investigation into frequency-dependent loudness variation as experienced in sound-field conditions (**Chapter 5**). Although loudness is not a direct cue for music listening, its relationship to pitch perception warrants investigation in a study of this kind.

Chapter 6 concludes this thesis with a brief general discussion of the study and concluding remarks, and offers recommendations for future work.

Chapter 2

FREQUENCY DISCRIMINATION ABILITIES OF COCHLEAR IMPLANT USERS IN SOUND FIELD

Pretorius, L.L. & Hanekom, J.J. (2008) Free field frequency discrimination abilities of cochlear implant users. *Hearing Research, 244,* 77-84

2.1 INTRODUCTION

CIs provide many profoundly deaf individuals worldwide with partially restored hearing ability. Although not quite comparable to NH, present multi-channel CI technology can transmit sufficient speech information to allow useful perception of speech (Wilson & Dorman 2007) and environmental sounds (Reed & Delhorne 2005). Music perception abilities, however, remain poor (Gfeller et al. 1997; Leal et al. 2003; Gfeller et al. 2005). Several music perception studies have shown that some cochlear implantees are able to perceive features such as tempo, rhythm (Gfeller & Lansing 1991; Kong et al. 2004) and lyrics (Looi et al. 2004; Gfeller et al. 2005) fairly well, but that perception of pitch-related features proves extremely challenging to most listeners (Gfeller & Lansing 1991; Gfeller et al. 1997; Kong et al. 2004; Sucher & McDermott 2007; Galvin et al. 2007). Even when pitch cues are presented in the absence of any confounding temporal cues, discrimination resolution of less than several semitones is not consistently observed in sound-field conditions (Pijl 1997).

The inability of otherwise NH listeners to perceive music successfully has been attributed to pitch perception deficits at various levels within the auditory processing system (Peretz 2002; Peretz & Hyde 2003; Foxton et al. 2004). Although both speech and music contain frequency information, fine spectral analysis appears to be more important for successful music perception than it does for speech perception. Adequate speech understanding can be achieved with as little as four spectral channels (Shannon et al. 1995), while at least 32 channels seem to be necessary for successful music perception (Smith et al. 2002; Kong et al. 2004). Shannon (2005) further argues that speech and music have different requirements for spectral resolution, which is in line with findings from studies regarding the neurocognition of speech and music in NH listeners (Zatorre et al. 2002).

Given the importance of pitch for successful music perception (in the Western tonal tradition), it follows that insufficient transmission of frequency information to the auditory system will adversely affect the quality of music perception. Despite pitch perception being mediated by both a place and a rate code (and possibly also a combination of the two) (Smith et al. 2002; Moore 2003) correct tonotopic representation appears to be especially important to successful pitch perception in NH listeners (Oxenham et al. 2004). CI-mediated transmission of frequency information also relies on the place pitch mechanism to a large extent (Swanson et al. 2009). After a signal has been passed through a set of bandpass filters, specific electrodes associated with these are activated, with the stimulus strength corresponding to the energy in the particular filter band. Spectral peak extracting algorithms (e.g. SPEAK and ACE) may then select the number of filters containing the strongest output and only apply stimuli to the associated electrodes. The electrodes in turn stimulate neural populations at (ideally) discrete positions along the length of the cochlea. However, because much of the frequency range relevant to human auditory tasks is handled by only a few electrodes during electrically mediated hearing, only limited spectral resolution is available to CI users. Correct transmission of frequency information is further hindered by channel interaction (Hughes & Abbas 2006) as well as mismatched frequency-to-place mapping (Fu & Shannon 2002), causing unintended neural populations to be stimulated.

The work reported in this chapter investigated the frequency discrimination abilities of cochlear implant users when presented with pure tone stimuli, across a wide frequency range, through their own processors in sound-field conditions. Because music is a multifaceted and acoustically complex auditory input, efforts to improve music perception abilities of CI users may benefit from studying music in terms of its comprising components, similar to the modular approach suggested for studying music-related deficits in neurologically impaired individuals (Peretz & Coltheart 2003).

With reference to the conceptual framework introduced in Figure 1.1 (Chapter 1), frequency discrimination represents an important module within the pitch processing stream during music perception. As shown in Figure 1.1, frequency discrimination contributes to the pitch analysis module at a processing level where the two primary dimensions of music information occur independently before combining at the next hierarchical level to effect "whole" melody perception. Assessing sound-field frequency discrimination abilities may thus present an early step in a modular approach towards music perception in CIs, since it would evaluate the system's ability to transmit one of the fundamental components of music in a setting which a user would perceptually be confronted with. By using simple stimuli, results of behavioural experiments can potentially be interpreted in terms of sound processing and associated neural stimulation, possibly providing a link between results from pitch perception studies performed with stimuli delivered at electrode level (Pijl & Schwarz 1995a; Pijl 1997) and those conducted at a perceptual level where stimuli consist of true music examples presented through users' speech processors (e.g. Gfeller et al. 2005).

The aim of the work presented in this chapter was twofold. Firstly, the typical frequency discrimination ability of CI users in sound field was investigated. At the time the investigation was initiated such data had not yet been measured systematically over an extensive frequency range and it was therefore deemed appropriate to measure. (Such data have in the meantime also been reported by other investigators (e.g. Gfeller et al. 2007) and the results of the investigation reported here are in line with those earlier ones.) Secondly, the investigation aimed to determine whether the observed frequency discrimination ability could be influenced by the stimulus frequency's position relative to filter response curves (see Figure 2.1). Several studies have shown that intermediate pitch percepts can be generated by dual electrode stimulation (McDermott & McKay 1994; McKay et al. 1996; Donaldson et al. 2005; Kwon & Van den Honert 2006). The reasoning behind the approach was that if the intermediate pitch percepts obtained from electrode level stimulation are linked to the differential distribution of current within a filter band, it should be possible to replicate the trend in sound field, since stimulus frequencies at different positions along the filter response curve would cause differential activation of more than one electrode quasi-simultaneously.

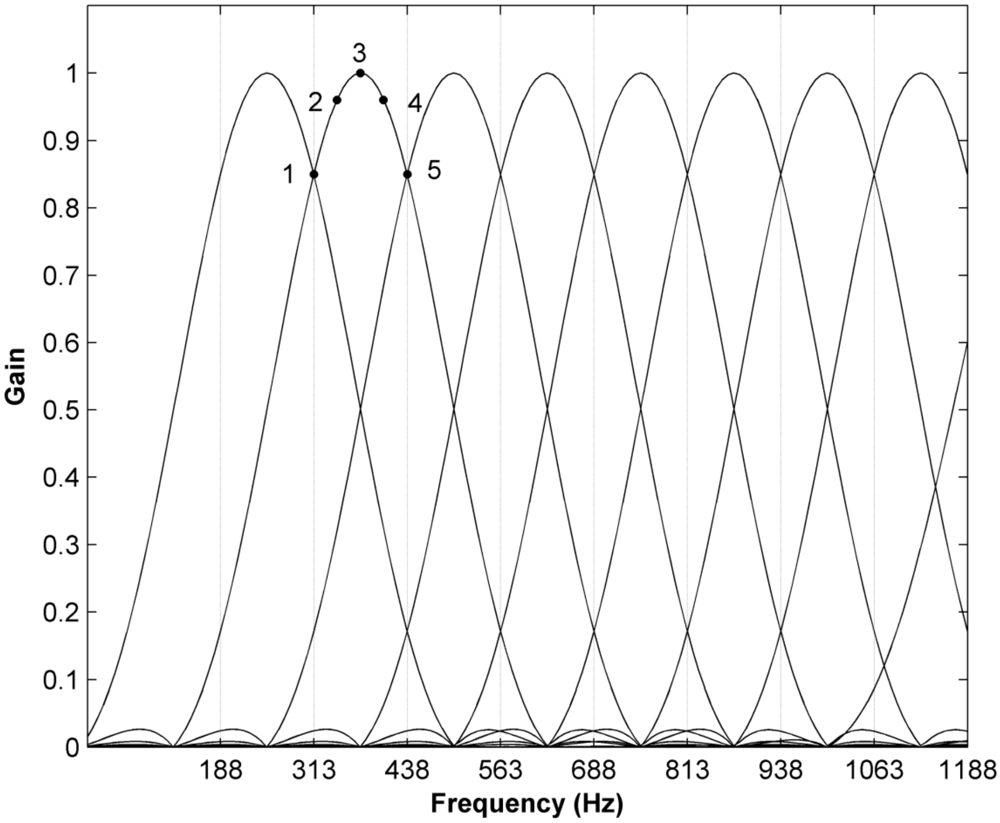

Figure 2.1: Frequency response associated with different filters as implemented in the ACE processing strategy. Numbers represent the different positions relative to which frequency discrimination thresholds were measured (indicated for one filter only).

2.2 METHOD

2.2.1 Subjects

Five post-lingually deafened adults implanted with the Nucleus 24 array participated in the investigation. All subjects had experienced profound hearing loss for more than 10 years and had more than six months' experience with the implant system. Subjects all used ACE as their normal speech processing strategy, with either the Esprit 3G or the Freedom speech processor, in a monopolar stimulation mode. Three participants were implanted bilaterally, but for these subjects only the implanted ear subjectively regarded by the subject as the clearer of the two was used during the course of the study. Electrodes 1 and 2 were inactive for S3, S10 and S14, while electrodes 1 and 22 were inactive for S13. All electrodes were active for S11. All subjects gave written informed consent, according to the guidelines from the relevant ethics committee, for their participation prior to commencement of the study and were compensated for their time at the end. Three subjects have participated in previous CI studies at our laboratory. The context in which the sample (which was one of convenience) was recruited is described in the Appendix and other relevant demographic details are provided in Tables A1 and A2 (Appendix A).

2.2.2 Stimuli

Since the study concerned assessment of CI users' frequency discrimination abilities in sound-field conditions, stimuli were chosen and presented to approach everyday listening conditions. Firstly, individualised reference frequency sets were compiled for each subject according to frequency responses of filters within the test range, as specified by the subject's clinical map. The test frequencies ranged between 250 and 1250 Hz, i.e. the frequency range with linearly spaced filters. This frequency range corresponds to fundamental frequencies of musical notes between C4 and D6.

All stimuli were 500 ms pure sine tones generated in Matlab 6.5[2] on a personal computer. Harmonics were suppressed by 45 dB or more across the frequency range used, as determined with a Larson-Davis 824 type 1 sound level meter. Amplitude ramps (30 ms) were included at the beginning and end of each tone to reduce onset clicks. Reference frequencies corresponded to frequencies at the peak of each filter, the cross-over between adjacent ones, and two points midway between the peak and cross-over positions on either side of the peak (see Figure 2.1). Thus, depending on the settings (filter width and active electrodes) specified by the subject’s clinical map, a set of between 23 and 27 reference frequencies could be compiled for each subject. A subject’s frequency discrimination ability was determined relative to each reference frequency within his/her individual set (also see Figure 2.2 for exact reference frequencies). Initial probe frequencies were always 100 Hz higher than the reference frequency, from where the difference between reference and probe frequencies was adapted according to listener response. Since the tested frequency range only spanned the region where filters are linearly spaced, using a fixed initial step size ensured that approximately equally spaced neural populations were targeted initially, regardless of reference frequency. More details regarding probe frequencies are provided in the section on experimental procedure.

Because stimuli were to be presented in sound field, it was important to present stimuli at the same perceived loudness level to all subjects, rather than at a fixed intensity level. A loudness estimation procedure was hence used to compile subject-specific sets of intensity levels corresponding to different loudness levels, by scaling 1 kHz tones ranging between the lowest and highest, yet comfortable, loudness levels. Stimuli were presented 20 times at each of 10 equidistant intensity levels within this range to derive an average estimated loudness percept at each level. All stimuli were subsequently presented at the intensity level corresponding to a subject’s 75% perceived loudness level, interpolated from the curve obtained with the 1 kHz tones. These were 80, 90, 80, 85 and 77 dB for subjects 3, 10, 11, 13 and 14 respectively.

Stimuli were not loudness balanced to one another. Although balancing is usually performed to ensure that loudness is not used as a cue (e.g. Collins & Throckmorton 2000) it was

---

[2] www.mathworks.com

reasoned that for the present experiments, electrode loudness balancing performed during the normal clinical mapping procedure would be sufficient (and preferred, as argued later) for sound-field testing conditions. After initial electrode-specific determination of threshold (T) and comfort (C) values, the clinical mapping procedure entails balancing the loudness of electrodes at C-level by adjusting C-values, and balancing the electrodes at 25% and 50% above T-level by adjusting T-values. Electrodes are subsequently compared in groups of five with an overlap of one electrode. The T- or C-level of a particular electrode can then be adjusted if it does not sound equally loud as its neighbours.

It was further reasoned that electrode loudness balancing would be adequate for the present experiments using pure tone stimuli, as these would primarily activate one electrode with weaker activation of neighbouring electrodes. Assumedly loudness-balanced electrodes should result in pure tone stimuli that are largely loudness balanced. To test the assumption, and whether frequency discrimination thresholds were influenced by using unbalanced stimuli, two subjects (S11 and S13) participated in an additional experiment to compare frequency discrimination abilities at a number of frequencies from the original range. Both loudness-balanced and unbalanced stimuli[3] were used. In the unbalanced condition electrodes were loudness balanced according to the usual clinical mapping procedure, but not the stimuli. Loudness balancing was performed by comparing sound-field loudness of pure tones at 12 evenly spaced frequencies between 200 and 1200 Hz, to that of a 1 kHz tone presented at the subjects’ respective 75% loudness levels. The listener’s task was to adjust probe stimuli to be equally loud, just louder and just softer than the reference stimulus. Three repetitions per frequency had to be completed per task, from which the average deviation from the 75% loudness level was calculated from all the data at a particular frequency.

The two subjects were always consistent in their adjustment of softer, louder and equally loud. Results showed that no more than a 4.7 dB (on average 1.8 dB) deviation from the reference 75% loudness level was necessary for equal sound-field loudness perceptions across the frequency range tested. Furthermore, as will be shown later, no significant differences in frequency discrimination thresholds were found for clinically loudness-

---

[3] This is defined as electrodes being loudness balanced according to the usual clinical mapping procedure, but not stimuli.

balanced maps (unbalanced condition) compared to those with loudness-balanced stimuli. Together, these results served as further motivation for not performing loudness balancing of stimuli in the main experiments. (The loudness balancing tasks performed here prompted a subsequent investigation regarding frequency-dependent loudness variation experienced by CI users in sound-field listening conditions, as described in Chapter 5.)

All stimuli were presented in a sound-proof booth. Standing waves produced in the sound booth may be a possible concern in sound-field listening conditions. Pure tones (10 s signal duration) were recorded across the frequency range and analysed. At all intensities used, the measured intensities were within 0.5 dB of the setpoint, with no measureable variation in the signal intensity, so that a possible influence of standing waves was ruled out. Total harmonic distortion was never larger than 3.5% (1.2% on average) across the entire range of intensities and frequencies used. The second harmonic was always the largest and was always suppressed by more than 35 dB. These measurements were taken with the same sound level meter as mentioned earlier.

The speaker (Yamaha MS101 II) was located approximately 1 m in front of the subject, but on the side of the implanted test ear. To ensure that the speech processor microphone did not produce significant harmonic components, harmonic distortion of the combination of loudspeaker and microphone was measured. Across the frequency range used, the largest harmonics were always suppressed by more than 35 dB.

2.2.3 Experimental procedure

Frequency discrimination thresholds in both the loudness-balanced and unbalanced conditions were determined using an adaptive two-alternative forced choice (2AFC) procedure. Each trial consisted of two 500 ms pure tones separated by a 1 s interstimulus gap, with one tone always corresponding to the reference frequency and the other being the probe frequency. Subjects were asked to indicate which of the two tones had the higher pitch by selecting either of two buttons indicated on the screen. No feedback was provided and subjects were not allowed to re-listen to a tone pair. However, a subsequent tone pair would not be presented before user response had been received, allowing subjects enough time to make a decision. Three practice trials, unknown to the subjects, were included at the start of each experiment session. These had to be completed successfully (i.e. correct response provided on three successive presentations) before the actual session commenced. Since a single session could be quite lengthy, an indicator showing estimated progress was included in the user interface.

An experiment session consisted of threshold determinations relative to three reference frequencies performed in triplicate. The three reference frequencies were preselected from the set of reference frequencies described earlier and were, as far as possible, of the same frequency class (i.e. peak, cross-over or midway frequencies relative to filter response curve). During threshold determinations reference frequencies were presented randomly to prevent the subject from tracking the change in probe frequency. For the same reason, reference and probe tones were also randomly assigned to either the first or the second presentation interval. The software controlling the experiment was, however, designed to keep track of subjects' previous responses at a specific reference frequency, in order to present stimuli correctly according to the adaptive technique. In order to reliably determine the average discrimination threshold relative to each reference frequency, two repetitions of each experiment session had to be completed, resulting in six threshold values at each reference frequency.

The first presentation of a probe frequency was always 100 Hz above the relevant reference frequency. Step size (difference between reference and probe frequency) was subsequently adjusted according to Levitt's (1971) transformed up-down staircase technique using a 2-down/1-up decision rule. A staircase technique allows the test stimulus to alternate between the values where it is just discernable from the reference stimulus and just not discernable, by tracking the listener's response behaviour. As soon as a non-discernable value is reached, the difference between probe and reference stimulus is increased until the probe can again confidently be distinguished from the reference. Each such response alternation is counted as a response reversal. A presentation run constitutes a succession of stimulus presentations between two response reversals. According to the decision rule used here, the difference between the probe and reference frequencies would only be reduced after two successive correct responses, while it would increase after a single incorrect response. This allowed the

71% correct point to be approached, which was accepted as the threshold. By adapting the step size according to a specific factor after successive responses, the discrimination threshold could be homed in on accurately. Step size changed by a factor of 2 (i.e. halved or doubled relative to the previous difference) during the first six response reversals, then by 1.7 for reversals 7 through 12 and by 1.3 for the last six reversals. An experiment session terminated once 18 response reversals had been recorded for each reference frequency. The threshold relative to each reference frequency was calculated as the average of the last five mid-run estimates.

Experimental sessions were scheduled according to subjects' availability, and hence data collection was completed over the course of several weeks. Subjects were briefed on the aim of the study and the nature of the experimental procedure prior to commencement, but were not aware of the algorithm controlling stimulus presentation.

2.3 RESULTS

Figure 2.2 shows the average normalised frequency discrimination thresholds obtained for the individual subjects relative to each reference frequency. Averages were calculated from six data points at each reference frequency. Because behavioural data obtained during psychophysical studies are prone to some variability, the Extreme Studentised Deviation (ESD) technique was used to test for outliers, which were then excluded from the average. No more than one outlying observation was found in cases where outliers were detected, leaving at least five data points from which to calculate the average.

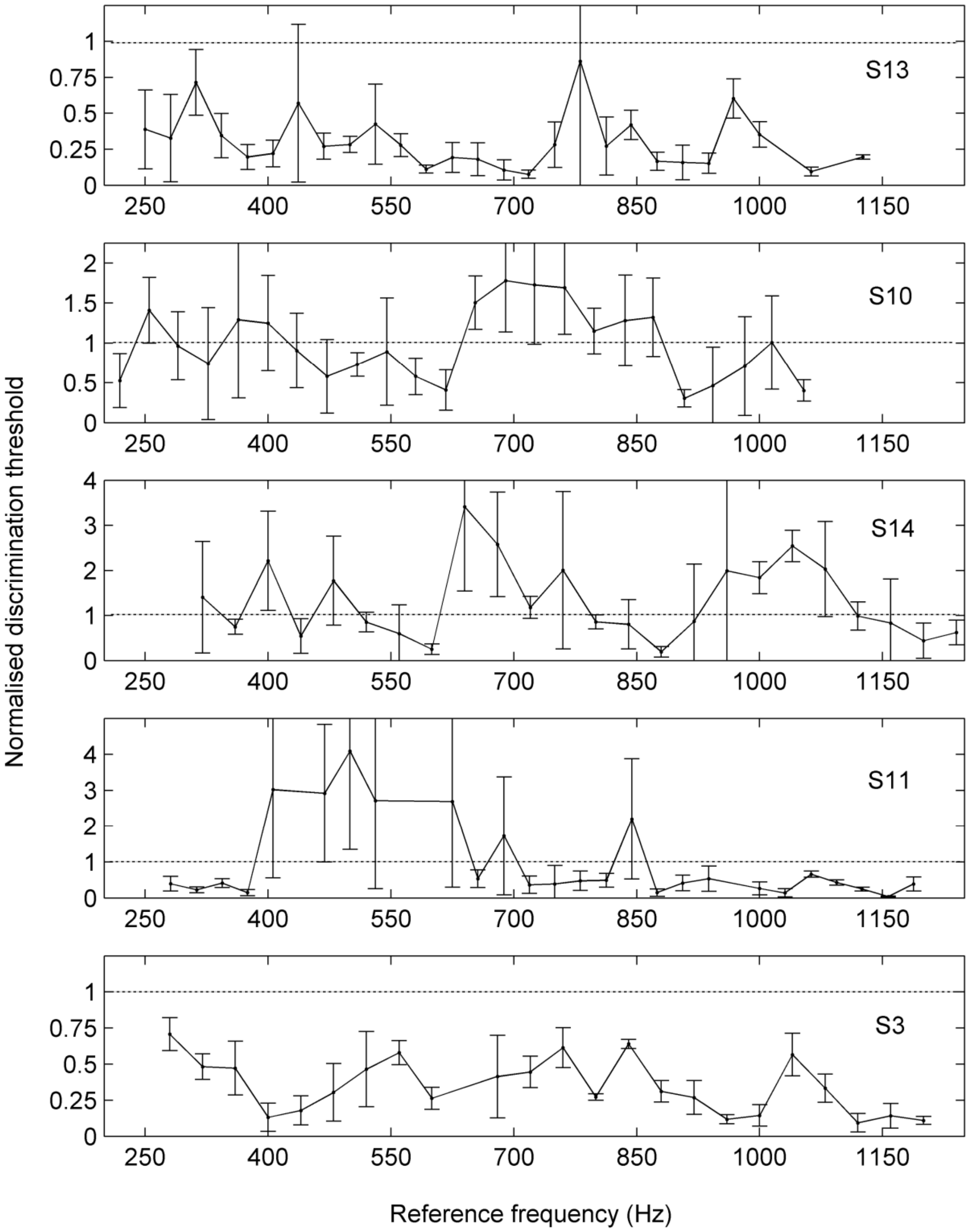

Figure 2.2: Average normalised frequency discrimination thresholds as a function of reference frequency. Thresholds are normalised to filter width as specified by the applicable frequency allocation table. The dashed line indicates one filter width. Individual subjects' results are shown in the respective panels.

Cases for which a frequency discrimination threshold could not be determined,[4] resulting in fewer than five replications for a specific reference frequency, were excluded from the analysis. Less than 3% of the total number of data points was excluded. Because of substantial intersubject variability, discrimination thresholds were normalised relative to filter width to allow all subjects' results to be viewed on the same scale. For the purposes of this study, a filter's width was defined as spanning the frequency range between the cross-over frequencies of two flanking filters. It should also be noted that because subjects used different frequency allocation tables, reference frequencies were not the same for all subjects. Figure 2.2 thus serves as a general view of the frequency discrimination trend across the frequency range tested. The dashed line indicates thresholds equal to one filter width.

It is evident from the data that not all subjects performed equally well during frequency discrimination tasks, nor was frequency discrimination behaviour similar across the frequency range tested. For two subjects, average discrimination thresholds were always less than one filter width, while results of the other three subjects showed discrimination thresholds of more than one filter width at several data points. For S11 large thresholds were consistently observed in the frequency region between 400 and 625 Hz, which may be attributed to subject-specific anatomical or physiological factors, since thresholds larger than those observed at other reference frequencies also presented in this region in two re-test sessions. However, all subjects seemed to exhibit a similar frequency discrimination trend in that some reference frequencies were associated with small discrimination thresholds, while others were associated with larger thresholds. Despite this observed variability, a large proportion of threshold values were smaller than one filter width, which a binomial analysis showed to be highly significant ($p < 0.001$). The analysis is essentially a hypothesis test whereby the probability associated with the mean not being equal to a forecasted value owing to a systematic influence of a variable is determined. The result indicates that these CI users were often able to discriminate pure tones less than one filter width apart in a sound-field situation.

---

[4] Frequency limits, above which the experiment would terminate, were set at initial experiment set-up. It sometimes happened that listeners failed to approach a frequency discrimination threshold, usually observed as very large step sizes. In such a case the specific run would terminate and it would be recorded in the results sheet that a frequency limit had been reached during the specific run. Frequency limits were always set at least 1000 Hz above the reference frequencies tested.

As mentioned earlier, minimal deviation from the 75% loudness level was necessary to achieve equally loud stimuli across the frequency range tested. Results from two subjects who repeated the frequency discrimination task with both balanced and unbalanced stimuli to test the influence of loudness imbalance are shown in Figure 2.3. Frequency discrimination thresholds smaller than one filter width were consistently observed with both balanced and unbalanced stimuli. For S11 the four (unbalanced) data points larger than one filter width presented in the same frequency region in which large frequency discrimination thresholds were observed during the initial discrimination tasks. Results of a Mann-Whitney test showed no statistically different frequency discrimination behaviour between balanced and unbalanced stimuli ($p > 0.05$).

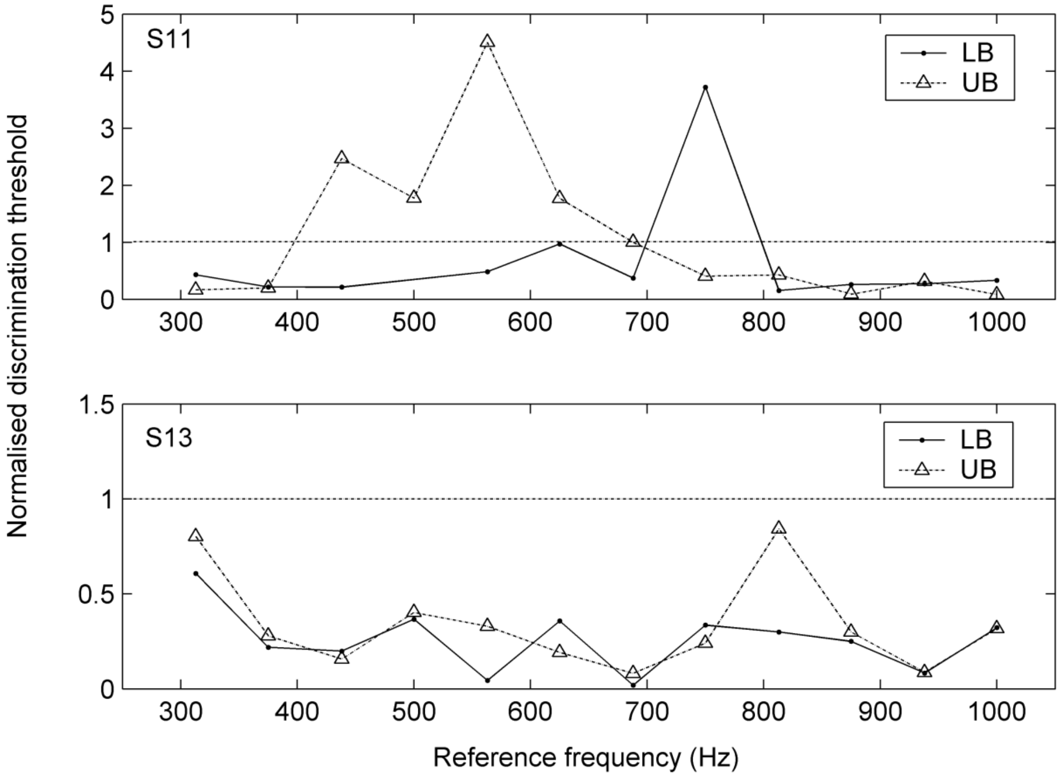

Figure 2.3: Average normalised frequency discrimination thresholds as obtained with loudness-balanced and unbalanced pure tone stimuli for two subjects.

In order to determine whether the frequency discrimination behaviour shown in Figure 2.2 could have been influenced by speech-processor-related parameters, the effect of filter location (a low, middle and high frequency region from the tested frequency range) as well as reference frequency position relative to the filter's frequency response curve were investigated. Figure 2.1 provides a schematic representation of the reference frequency positions and filter locations. Positions 1 and 5 correspond to the left and right cross-over frequencies respectively, with position 3 associated with the peak frequency. Positions 2 and 4 are frequencies halfway between the peak and left and right cross-over frequencies respectively. Owing to different absolute reference frequencies, as well as some missing value cases (as explained earlier) the data set had to be reduced somewhat to obtain a complete set for analysis purposes.

Figure 2.4 shows average normalised thresholds at five positions along the filter response curve for filters located at a low, middle and high frequency (relative to the tested range) for each subject. Results from a binomial test using the complete data set showed that a statistically significant proportion (73%, $p < 0.05$) of discrimination thresholds was smaller than one filter width. Furthermore, 64% of thresholds were smaller than half a filter width at position 3 ($p < 0.05$). This compares to significant proportions (78%, 71%, 67% and 67%, $p < 0.05$) of discrimination thresholds at positions 1, 2, 4 and 5 being smaller than one filter width, but not smaller than half a filter width. The result shows that these listeners were often able to reliably discriminate different pitches falling within the same filter band, especially when the reference frequency corresponded to that of the filter's peak (position 3).

A general linear model with a three-factor design further showed that both filter location and position of the reference frequency relative to a filter's frequency response curve, as well as their interaction, have significant effects on frequency discrimination behaviour ($p < 0.02$). However, further analyses indicated that differences with regard to frequency discrimination behaviour were not always the same for a specific position, but rather that the position effect appeared to be dependent on the particular filter.

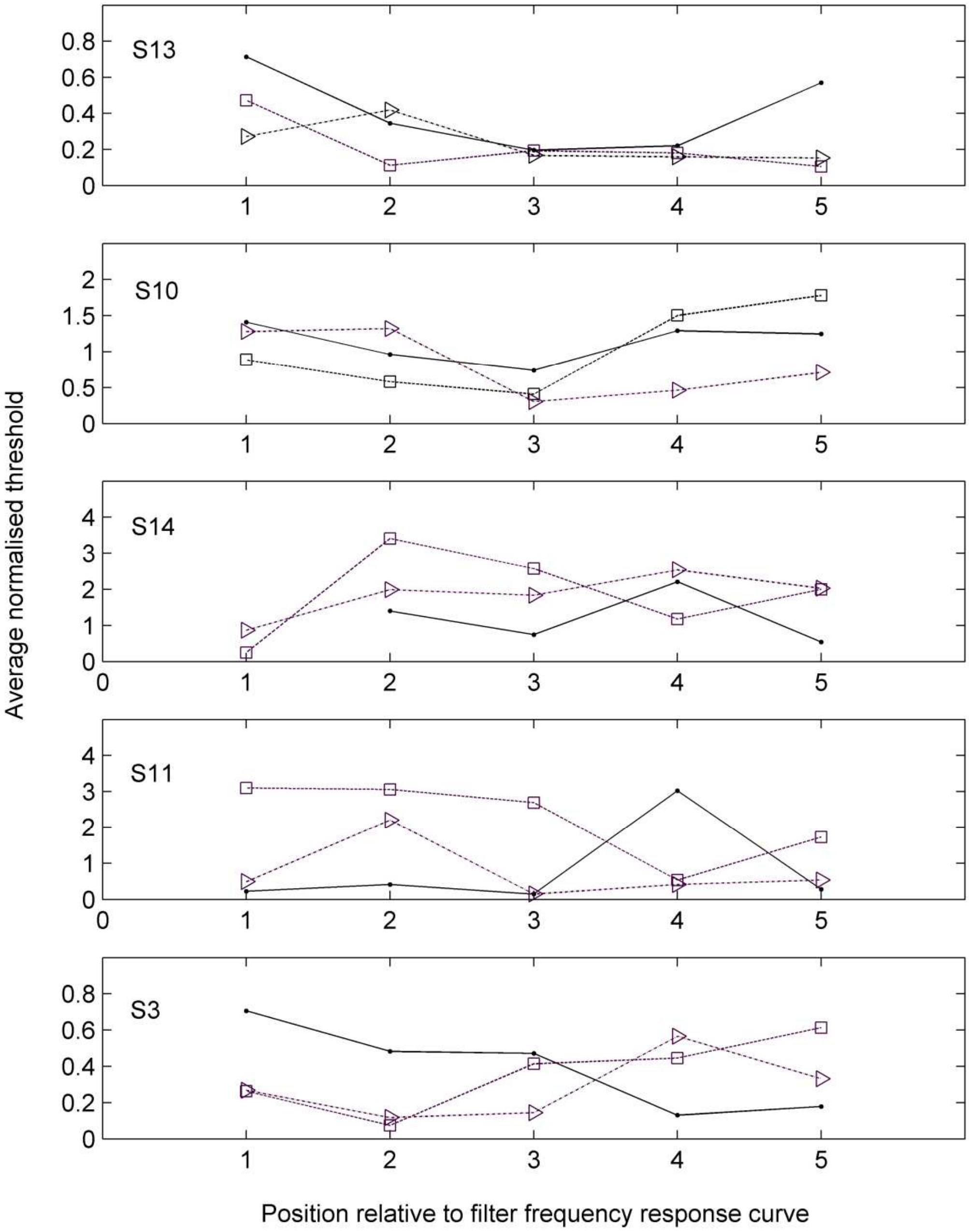

Figure 2.4: Average normalised frequency discrimination thresholds as a function of position along filter frequency response curve. Different lines represent frequency discrimination behaviour from filters representing low (dots), middle (open squares) and high (open triangles) frequencies in the range tested. Individual subjects' results are shown in the respective panels.

2.4 DISCUSSION

The results from this investigation show that (i) CI users were able to discriminate between different pure tone frequencies falling within one filter band and (ii) that discrimination behaviour was influenced by the position of reference frequency relative to the filter's frequency response curve. Taken together, the results indicate that finer frequency resolution than would be expected, could be available to CI users.

This is an unexpected result, considering the dominant place pitch effect in electrically mediated hearing (McDermott & McKay 1997) and, given fixed stimulation rate, the limited temporal pitch cues available in the present experimental conditions. However, earlier pitch perception studies conducted at the electrode level have shown that dual stimulation of adjacent electrodes, either simultaneously (Townshend et al. 1987; Donaldson et al. 2005; Firszt et al. 2007) or separated by short gaps (McDermott & McKay 1994; Kwon & Van den Honert 2006) may generate more pitch percepts than dictated by the number of electrodes. Firszt et al. (2007) showed that the effect could be associated with the proportion of current delivered to two simultaneously activated electrodes, so that different pitch percepts could effectively be created by deliberate manipulation of current on adjacent electrodes. This is in line with results from studies regarding vowel discrimination where Dorman et al. (1992; 1996) suggested that the frequency of vowel formants could be signalled by the energy balance between two adjacent filter bands.

The Nucleus device does not allow true simultaneous dual electrode activation and therefore improving frequency resolution through current manipulation may not be feasible. However, quasi-simultaneous electrode activation patterns can activate different neural populations within the same stimulation cycle. The neural activation strength is strongly affected by the stimulating frequency's position relative to the filter response curve. Frequency input at different positions along the filter response curve may thus naturally create a neural targeting mechanism by causing small shifts in the activated neural population. It is possible that the smaller than expected frequency discrimination thresholds observed here may be attributed to operation of such a mechanism.

Each position along the filter response curve investigated in the present investigation represented a frequency that falls within two or three adjacent filter bands. As shown in Figure 2.1, no more than 6 dB attenuation (gain of 0.5) will be experienced at any of these positions. This should be sufficient to contribute substantially towards activation of these filters' associated electrodes, as this would normally be within the dynamic range of an electrode. Such differential activation of neighbouring electrodes may result in slight shifts in the activated neural population upon slight adjustment of the pure tone input frequency. This would in turn cause an overlapping neural population to receive substantial activation from more than one electrode within the same stimulation cycle, thereby increasing the strength of activation in the specific neural population. Given that stronger activation increases action potential frequency, stronger activation in a specific neural population might be interpreted as a more salient frequency cue. A schematic explanation of the mechanism is offered in Figure 2.5, showing how filter output is associated with electrode activation and consequently drives neural activation. As can be seen in the schematic representation, a frequency that corresponds to the peak of the frequency response curve would activate three electrodes simultaneously (represented by the red, green and blue ranges), and consequently three overlapping neural populations. The neural population receiving strongest activation (i.e. most overlap) may thus generate a more salient frequency message that feeds into subsequent processing modules.

It is interesting to note that in all the aforementioned studies, improved frequency resolution was observed when stimuli were manipulated deliberately at the outset of the experiment so that different current proportions would be delivered to adjacent filter bands. Here, however, stimuli were presented in sound field and electrode activation was governed completely by speech processor output. This suggests that proportional current delivery as determined by the speech processor may translate effectively to differential neural activation. A differential neural activation mechanism as might be in play in these experimental conditions may therefore have a similar perceptual effect as deliberate current steering. It further appears that the effect is strong enough to translate to a perceptual outcome even when it functions within the complete, electrically stimulated peripheral auditory system. Although the number of additional pitch steps that can be made available through such a mechanism was not specifically determined in this study, the observation of a statistically significant proportion of discrimination thresholds being smaller than half a filter width at position 3 implies that at least double the present frequency resolution may be available at some positions. It thus appears that frequency resolution may be improved by exploiting the characteristics of existing speech processors. It should be noted, however, that physiological and anatomical characteristics, such as neural survival, position of electrode array relative to the basilar membrane and formation of scar tissue may influence individual performances and the benefit derived from a possible functionally available endogenous steering mechanism as described here.

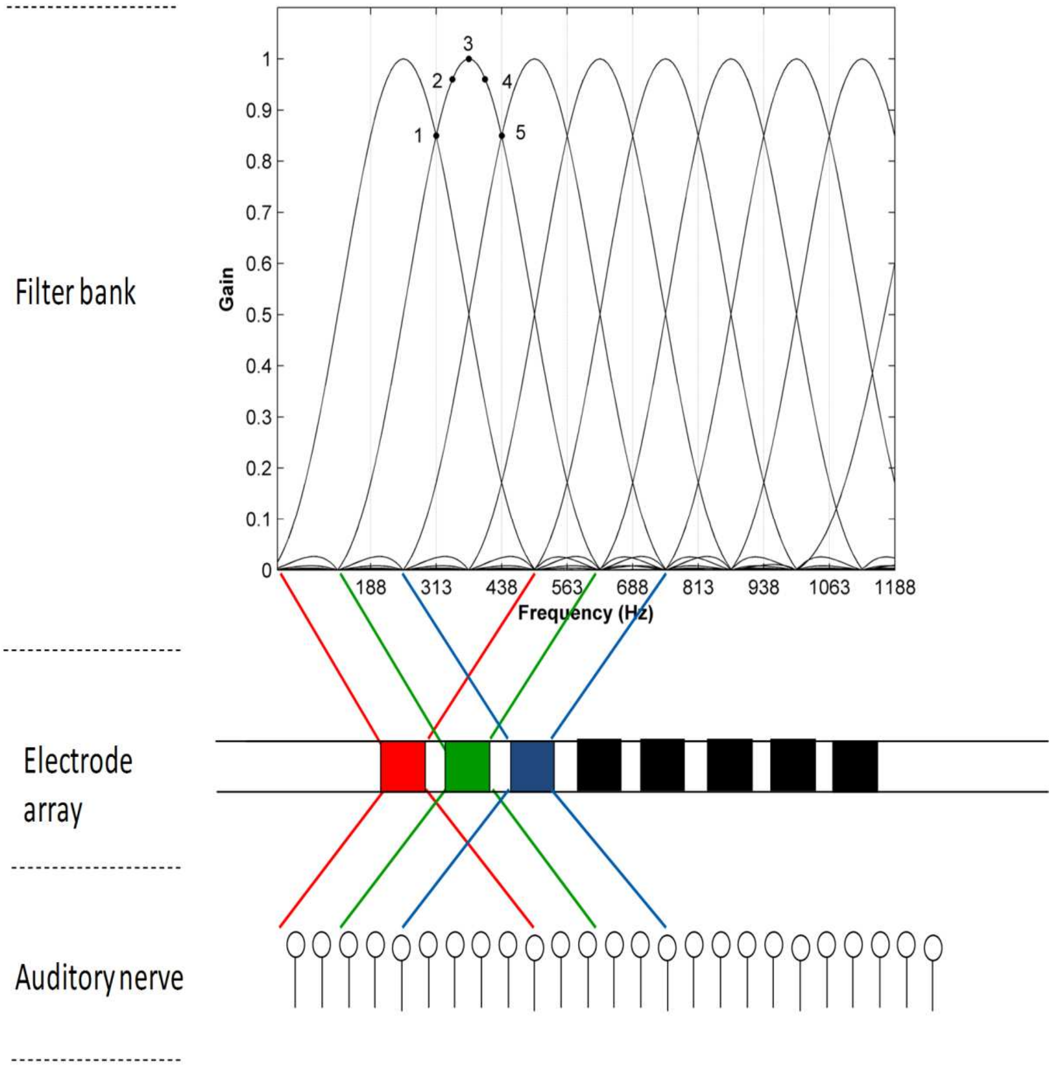

Figure 2.5: Schematic representation of possible mechanism of the neural steering mechanism as may be in operation during sound-field frequency discrimination

The frequency discrimination behaviour observed in this investigation also hints at CI users' possible music perception abilities. Although musical tones are rarely simple sinusoidal tones as used in this study, the results may be a relevant first indication of the level of pitch resolution that may be available to CI users when listening to music.

In Figure 2.6 frequency discrimination is expressed as the ratio of the actual threshold frequency to the reference frequency. This translates to the discrimination threshold in number of semitones, as shown on the right-hand ordinate.

Successful melody perception, regarded as central to music perception in the Western tonal style, relies strongly on the correct recognition of such relative rather than absolute pitch distance between two successive tones (Peretz 1990; Peretz & Zatorre 2005). The frequency of two tones an octave (12 semitones) apart will always be related in a 1:2 ratio, while two tones one semitone apart stand in a 1:1.06 ratio. Although the absolute frequency difference between two tones an octave apart increases exponentially from the lower to the upper end of the musical spectrum, the frequency ratio between them always remains constant.

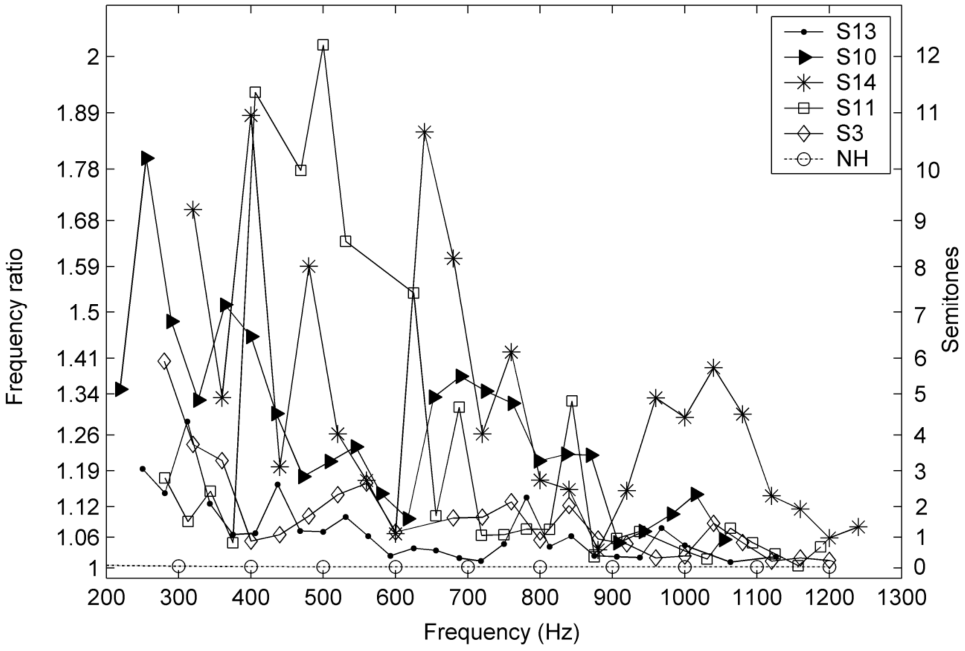

Figure 2.6: Ratio of absolute threshold frequency to reference frequency plotted as a function of reference frequency. The number of semitones associated with each ratio is shown on the right-hand ordinate. The dashed line represents frequency discrimination behaviour of NH listeners across the same range (Zwicker & Fastl 1999).

It follows then that to perceive such melodic contours successfully, pitch distances of at least one semitone should be recognised consistently across the entire musically relevant frequency range. This is indeed the case for NH listeners (Zwicker & Fastl 1999). However, as shown in Figure 2.6, CI listeners in this investigation displayed marked variation in their frequency discrimination ability across the range of base frequencies, with discrimination thresholds of the sample as a whole not consistently approaching one semitone across the frequency range tested. There is, however, a general trend of the ratio between frequency difference limen and reference frequency, and hence the number of semitones that can be discriminated, to decrease for CI users across the frequency range tested. Despite this being a direct consequence of filters being spaced several semitones apart at lower frequencies, while only a few semitones apart at higher frequencies, it does not change the implication that frequency differences between two tones may not be discriminated equally well in low and high registers. This observation is in line with findings that CI users achieve somewhat better sound-field melodic contour recognition of tone sequences toward the upper end of the musical spectrum (Galvin et al. 2007; Singh et al. 2009). Such inconsistent recognition of pitch distances may contribute to unsatisfactory music perception and subsequent appreciation.

Finally, it is necessary to consider the influence of loudness balancing on frequency discrimination thresholds. More important than loudness balancing between electrodes (as achieved in clinical mapping), would be loudness balancing between stimuli (which was not done in the main experiments of this investigation). A particular pure-tone stimulus activated not only a single electrode, but also adjacent electrodes as can be inferred from Figure 2.1. Therefore, some criticism may be expressed about not performing loudness balancing of stimuli in the present investigation.

Results suggest, however, that loudness balancing was not necessary for the particular task. Not only did stimulus loudness have to be adjusted only minimally from the unbalanced condition to achieve loudness balance, but results also showed no statistically significant difference in frequency discrimination behaviour for balanced and unbalanced stimuli. This suggests that sufficient loudness balancing was achieved in the clinical mapping procedure so as not to influence frequency discrimination behaviour significantly in sound-field conditions. This is an important result, considering that the objective was specifically to assess implant users' behaviour in conditions similar to those they would be confronted with during everyday listening using their standard maps. It was argued that, if a sound-field frequency discrimination test were to become part of a music perception test for cochlear implantees, it would be inappropriate to perform loudness balancing before commencement of such a test, as one would wish to measure music perception ability with a particular speech processor and a particular clinical map.

Furthermore, since frequency discrimination thresholds were often smaller than one filter width, the reference and probe stimuli would in these cases have activated the same electrodes, with loudness effects unlikely to influence frequency discrimination behaviour. This is borne out by the data of S13 shown in Figure 2.3. However, frequency discrimination thresholds larger than one filter width obtained for S11 between 400 and 625 Hz suggest that frequency discrimination data may have been influenced by loudness effects to some extent. Since these larger discrimination thresholds were repeatedly observed only in an isolated region during presentation of unbalanced stimuli, it is possible that an existing physiological or anatomical constraint, together with loudness effects, may have resulted in confusion. Possible loudness effects in these regions appear to be localised and subject-specific, and therefore do not influence the general conclusions of this investigation.

2.5 CONCLUSION

Sound-field frequency discrimination abilities of CI implant users appear better than expected. Although discrimination resolution is not equally good at all reference frequencies across the range investigated, it seems that two pure tones falling within the same filter band can often be discriminated successfully. The findings may be linked to the distribution of energy across adjacent filter bands, resulting in differential electrode and subsequently differential neural activation patterns. Attempts to improve CI users' music perception abilities may benefit from further investigation into exploiting existing sound processing features to improve spectral resolution, which in turn may contribute to better music perception for CI users.

Chapter 3

RHYTHM PERCEPTION BY COCHLEAR IMPLANTEES IN CONDITIONS OF VARYING PITCH

Pretorius, L.L, Hanekom, J.J. & Venter, P.J. (2010) Rhythm perception by cochlear implantees in conditions of varying pitch (in review, *Cochlear Implants International*)

3.1 INTRODUCTION

A melody can be described as a succession of tonal events that unfold over time, with individual events related to one another according to a tonal and temporal structure. This implies that pitch and rhythm information both contribute to constituting a perceptually meaningful melody. Moreover, pitch and rhythm patterns seem to function in an interdependent (Kidd et al. 1984; Boltz 1999; Lebrun-Guillaud & Tillman 2007) and often complementary (Jones 1993) way to create musically coherent wholes for perception. This is supported by the modular approach to music processing, which, based on results from neurocognition and lesion studies, suggests that separate yet parallel cortical processing pathways exist for time- and pitch-related information (Peretz & Coltheart 2003). The pathways are believed to be organised in a serial, hierarchical manner, with regard to both structure and function, which allows processing output from the two domains during earlier stages to combine at a higher hierarchical level to generate a coherent musical percept (Peretz & Kolinsky 1993; Schuppert et al. 2000; Peretz & Coltheart 2003; Griffiths 2003; Semple & Scott 2003).

Successful melody perception therefore requires adequate perception of pitch as well as rhythm information. It follows that constrained information transfer of either (or both) of these dimensions will hamper melody perception. Temporal resolution afforded by a CI, as inferred from gap detection thresholds, appears to be similar to that of NH listeners (Moore & Glasberg 1988; Shannon 1989; Grose & Buss 2007). These authors concluded that temporal resolution is probably determined at a post-cochlear stage, whereas pitch resolution is determined already at the cochlear level. It is thus reasonable to expect that CI users should perform better on melodic rhythm than on pitch perception tasks, and possibly even at a level comparable to NH listeners, since temporal information is probably transmitted to the auditory cortex relatively unconstrained.

However, studies regarding implant users' music perception abilities (e.g. Gfeller & Lansing 1991; Gfeller et al. 1997; Kong et al. 2004; Galvin et al. 2007; Cooper et al. 2008; Looi et al. 2008) have shown that although significantly better than pitch perception, rhythm perception is not unambiguous. Using pairs of simple monotonic rhythmic patterns, which subjects had to judge as the same or different, Gfeller et al. (1997) showed that CI users performed similarly to NH listeners. In the same study, however, when a discrimination judgement had to be made for more complex rhythmic patterns, implant users performed significantly poorer than NH listeners. Similarly, Kong et al. (2004) found, albeit in a small sample, that CI users performed 5–25% worse than an NH control group on a rhythm identification task.

Observations that simple familiar melodies are better recognised when pitch as well as rhythmic cues are available (Gfeller et al. 2002; Galvin et al. 2007) highlight the important contribution of rhythmic cues to CI-mediated melody perception. It is possible that when confronted with limited pitch information during melody perception, the auditory system relies on rhythm information to supplement insufficient information to achieve the best possible melody perception in constrained conditions. In this regard recent findings regarding the rhythm perception abilities of amusic listeners are important. Amusic listeners often find it difficult to follow pitch changes in a melody, which possibly stems from deficits in processing pitch information (Foxton et al. 2004). Within the framework of early independent processing of pitch and rhythm information (Peretz 1990; Peretz & Coltheart 2003) one would expect that rhythm perception should be unaffected by pitch processing deficits. However, Foxton et al. (2006) showed that amusic listeners' perception of pitch-varying rhythmic patterns was significantly poorer than that of NH listeners. The authors suggested that when pitch and rhythm information covary in the same sound sequence, the information is encoded as a unified representation rather than activating two separate processing systems. Insufficient pitch processing hence appears to interfere with rhythm perception.

CI users also experience pitch processing difficulties, although not derived from central processing impairments (see for example Koelsch et al. 2004) but rather owing to insufficient delivery of pitch-related information to the central auditory system (Rubinstein 2004; Sucher & McDermott 2007), and as such can be regarded as functionally tone-deaf. This raises the question whether implant users' rhythm perception ability is subject to the same constraints as those experienced by amusic listeners. If so, it could mean that already challenging melody perception, due to insufficient transmission of pitch information, may be complicated even further. In this regard it is important to keep the schema outlined in Figure 1.1 (Chapter 1) in mind. Rhythm perception represents processing of musical information features at an early hierarchical level from where output is sent forward to a subsequent hierarchical level to contribute to "whole" melody perception. However, based on results regarding amusic listeners' rhythm perception abilities, it is important to consider the possible influence of pitch processing deficits on rhythm perception, despite the two dimensions' processing being posited as largely independent at this hierarchical level. Thus according to the hierarchical processing framework as proposed in Figure 1.1, investigating rhythm perception at a perceptually early level while considering a possible cross-influence from the pitch dimension, could contribute to perceptual outcome at later levels being better interpreted.

In previous studies specifically aimed at investigating rhythm perception of CI users (e.g. Gfeller et al. 1997; Kong et al. 2004; Looi et al. 2008), pitch was not varied. However, covarying pitch and rhythm cues in a controlled manner may provide improved understanding of CI users’ rhythm perception ability with regard to music-like tone sequences. The work presented in this chapter investigated CI users’ rhythm perception for sound-field tone sequences of varying rhythmic complexity, in either monotonic (pitch-constant) or polytonic (pitch-varying) conditions. The objectives were (i) to determine what the typical rhythm perception ability of CI users is for simple sound-field tone sequences, and (ii) to assess whether pitch processing deficits influence rhythm perception to the same extent as for amusic listeners. Interpreting findings in the context of the neurocognition of music may lead to a better understanding of some of the underlying information processing that contributes to CI-mediated melody perception.

3.2 METHODS

3.2.1 Subjects

Seven post-lingually deafened adult users of the Nucleus device (24-electrode array) participated in the study. All subjects had experienced profound hearing loss for more than 10 years and had more than a year’s experience with the CI system. Except for one Sprint and one Esprit 3G user, all subjects used a Freedom processor. All but one of the CI listeners used the ACE strategy. Four subjects were implanted bilaterally, but used only the ear subjectively regarded as allowing clearer perception during the investigation in order to create a homogenous monaural listening test group.[5] None of the subjects had advanced musical training before onset of deafness, although S3 was a member of a church choir as teenager and S11 often attends folk music recitals. Other relevant detail is provided in Tables A1 and A2 (Appendix A).

---

[5] S11 used the right ear during this investigation (compare to Chapter 2). This was because she had been fitted with new maps on both implants shortly before this investigation commenced and found the right ear to give a more “natural” sound.

Seven age-matched NH control listeners also participated in the study. Normal hearing was defined as achieving audiometric thresholds of 30 dB HL or better at six octave frequencies from 250 to 8000 Hz. Only one NH listener had formal music training, but all other participants indicated that they enjoyed listening to Western tonal music and could informally participate in singing. All but one of the control listeners had previously participated in research in our laboratory.

As required by the relevant ethics committee, all subjects gave written informed consent for their participation prior to commencement of the study. Subjects were compensated for their time on completion of the tasks.

3.2.2 Stimuli

Stimuli consisted of pairs of simple tone sequences generated in Matlab 6.5 on a personal computer. Each sequence comprised five 100 ms pure sine tones separated by four brief intertone intervals (ITIs). Sequences were strung together prior to each presentation round to prevent the duration of the ITIs from being influenced by processing. Amplitude ramps (30 ms) were included at the beginning and end of each tone to reduce onset clicks. During each presentation round one sequence (probe) contained a longer ITI presented at a specific position, while the other (reference) did not. The longer ITI was presented either between the second and third, or fourth and fifth tones. The probe and reference sequences were separated by a 1500 ms gap.

A combination of rhythmic and pitch pattern complexity (two levels each) created four conditions of varying perceptual demand. Rhythmic complexity was determined by the ITI, which, together with the tone duration, determined the inter-onset interval (IOI). Gauging responses to IOI changes is a useful indicator of rhythm perception, since the rhythm of a sequence is primarily determined by the time between tones, rather than the duration of the tones themselves (Krumhansl 2000). Since tone duration was always 100 ms, the ITI was altered to create different IOI patterns. An isochronous sequence with a constant ITI of 300 ms represented the simple rhythmic condition, while an anisochronous pattern with ITIs alternating between 300 ms and 600 ms represented the complex rhythmic pattern. Base ITI in the isochronous condition was 300 ms, while 600 ms in the anisochronous condition. In the monotonic pitch condition all tones were presented at 500 Hz. The polytonic pitch condition involved varying pitch patterns, with the first tone always presented at 500 Hz and subsequent tones varying between 200 Hz and 2000 Hz at 100 Hz increments to ensure that pitch differences would be clearly audible to CI listeners. The frequency of successive tones could change in any direction (up, down or not at all), but never differed by more than 200 Hz from the preceding tone. The stimuli design strongly resembles that used by Foxton et al. (2006). Figure 3.1 provides a schematic representation of stimuli specifics.

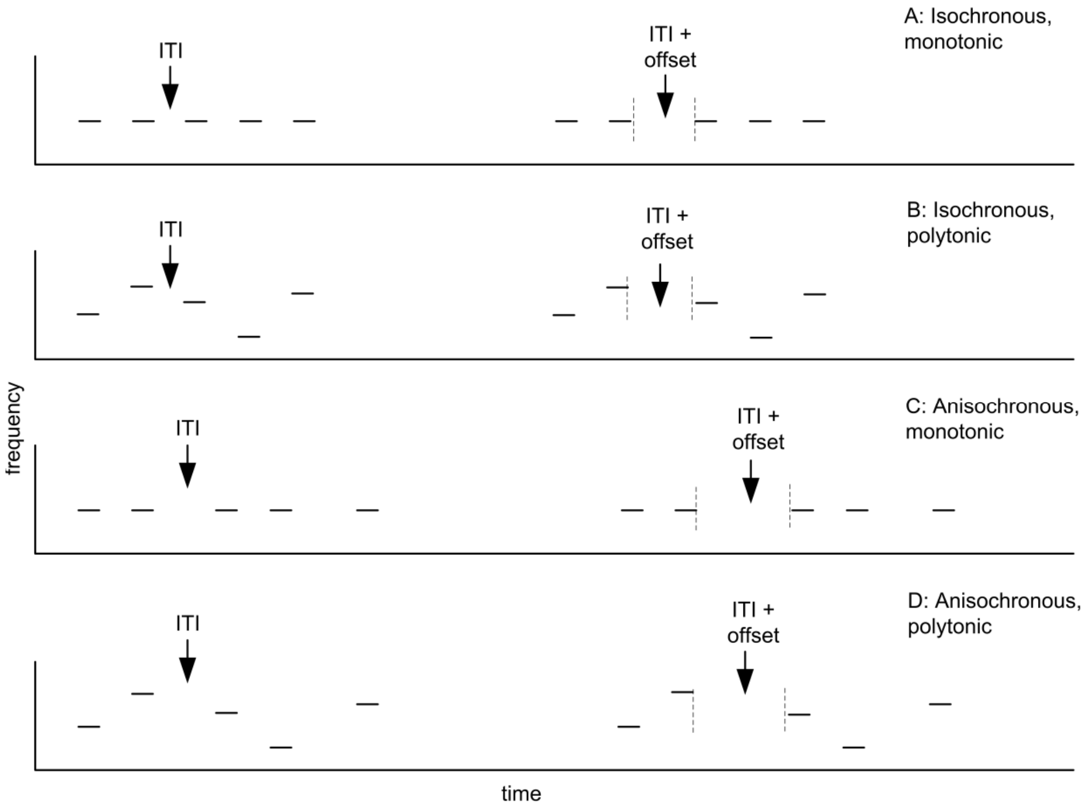

Figure 3.1: Schematic representation of test conditions. IOI = Inter-onset interval; ITI = intertone interval. Because tone duration always remained constant, ITI was adjusted to create different IOI patterns. Reference ITI duration was 300 ms in the isochronous conditions (A and B) and 600 ms in the anisochronous conditions (C and D).

All stimuli were presented in sound field. Subjects used their clinically assigned processors and settings. Stimuli were presented through a Yamaha MS101 II loudspeaker, which was placed approximately 1 m away, in front of the listener on the side of the test ear. Because of the sound-field set-up it was important to present stimuli at the same subjective loudness to each participant, rather than at a fixed intensity level. A loudness estimation task was performed prior to the rhythm perception tasks. Each listener's 75% perceived loudness level was determined for a 1 kHz pure sine tone, to which intensities of all other tone frequencies were subsequently loudness balanced.

Loudness balancing was performed to rule out any confounding loudness cues, especially considering the pitch-varying nature of the polytonic conditions. Frequencies selected at 100 Hz increments across the range to be used during the rhythm perception tasks were loudness balanced relative to a 1 kHz tone at the intensity earlier determined to represent the 75% perceived loudness level for each listener. The balancing task required listeners to perform three comparisons: adjusting the probe stimulus to sound equally loud, just louder and just softer compared to the reference stimulus. Each frequency was randomly presented three times during the course of the respective loudness adjustment tasks. The average deviation from the reference intensity was calculated from these nine samples and presentation intensity for frequencies selected for rhythm perception stimuli were stored in a look-up table for later use.

3.2.3 Experimental procedure

Rhythm perception ability was assessed in terms of ITI duration discrimination, using an adaptive 2AFC procedure. The probe sequence, randomly assigned to either of the two presentation intervals, was identical to the reference sequence, barring target ITI duration. The target ITI was also randomly presented between the second and third, or fourth and fifth tones, requiring subjects to listen to the whole sequence rather than focus on a specific ITI position. Together these measures reduced the possibility of listeners tracking the pattern of duration offsets, which could have introduced bias into the results. However, an equal number of response reversals had to be completed for each of the two target ITI positions, and results were analysed separately for the two positions.

The target ITI duration offset (the difference between the reference and probe ITI) was adapted according to subject response. In both isochronous and anisochronous conditions the initial ITI offset was 600 ms to allow subjects to hear the difference in ITI duration clearly. Thus for the isochronous condition, total target ITI amounted to 900 ms; for the anisochronous condition, the total intial target ITI duration was 1200 ms (600 ms base + 600 ms offset). A transformed up-down staircase procedure (2-down/1-up decision rule) was used to adjust ITI duration offset, converging on 71% correct (Levitt 1971). A test run concluded after 15 response reversals. For the first five reversals, the ITI duration offset was adjusted by a factor of 2; for the second five and last five reversals, the ITI duration offset was adjusted by a factor of 1.7 and 1.3 respectively. The average of the last five reversals was calculated as the average ITI duration threshold for each test run.

Each combination of rhythmic and pitch complexity represented a test condition and was tested six times (three sets of two test runs for each combination) with CI listeners. Since NH listeners’ results showed less variation, they completed only three repetitions of each test condition. Depending on subjects’ available time, more than one test condition could be tested during a single test session. Subjects were told that the target (lengthened) ITI could be between either the second and third or the fourth and fifth tones, and that the probe sequence could randomly be presented as either the first or the second presentation interval. Rhythmic patterns were also tapped out to subjects prior to commencement of the study. Subjects neither received feedback regarding their responses nor were allowed to re-listen to stimuli. Test trials were self-paced and a new stimulus was not presented until subjects had responded to the previous stimulus. Unbeknown to subjects, three practice trials were included at the beginning of each run to familiarise listeners with the stimuli and test procedure. Subjects were required to complete the practice trials successfully before actual testing was begun. Since an experiment run could last up to 20 minutes, a progress indicator was shown on the computer screen to keep listeners motivated.

3.3 RESULTS

Figure 3.2 shows mean individual ITI duration discrimination thresholds, as well as results averaged over all participants in a group, for the four experimental conditions. The top and bottom panels depict results for the target interval placed between the second and third, and fourth and fifth tones respectively. Individual subjects’ results represent the average of six repetitions (or three, in the case of NH listeners) per stimulus condition.

Since subjective response behaviour as measured in this study can be prone to between-session variability, each subject’s respective data sets were inspected for outliers using the ESD technique. Only three outliers were found among the 336 CI listener responses (six repetitions x four conditions x two target positions x seven subjects), in three different data sets. The reported average thresholds for those sets were thus calculated from the remaining repetitions. No outliers were found among the 168 NH listener responses (three repetitions x four conditions x two target positions x seven subjects). Rhythmic complexity (isochronicity or anisochronicity), pitch complexity (monotonic or polytonic) and position of target interval (between second and third, or fourth and fifth tones) were defined as within-subject variables, while hearing ability (CI or NH) was defined as the between-subject variable in a three-way repeated measures analysis of variance

CI and NH listeners performed similarly in all conditions ($F(1,12) = 1.62$, $p = 0.23$), indicating that the CI device allows uncompromised perception of the rhythmic information presented in this investigation. The finding is supported by the observation that none of the main effects produced a statistically significant interaction with hearing ability.

The influence of additional rhythmic and pitch complexity are noteworthy. Manipulating rhythmic complexity had a pronounced effect on rhythm perception ability for both CI and NH listeners ($F(1,12) = 89.77$, $p < 0.001$), as seen from discrimination thresholds that were three to four times higher for anisochronous than isochronous conditions. Manipulating pitch complexity, however, exerted less influence on rhythm perception ability.

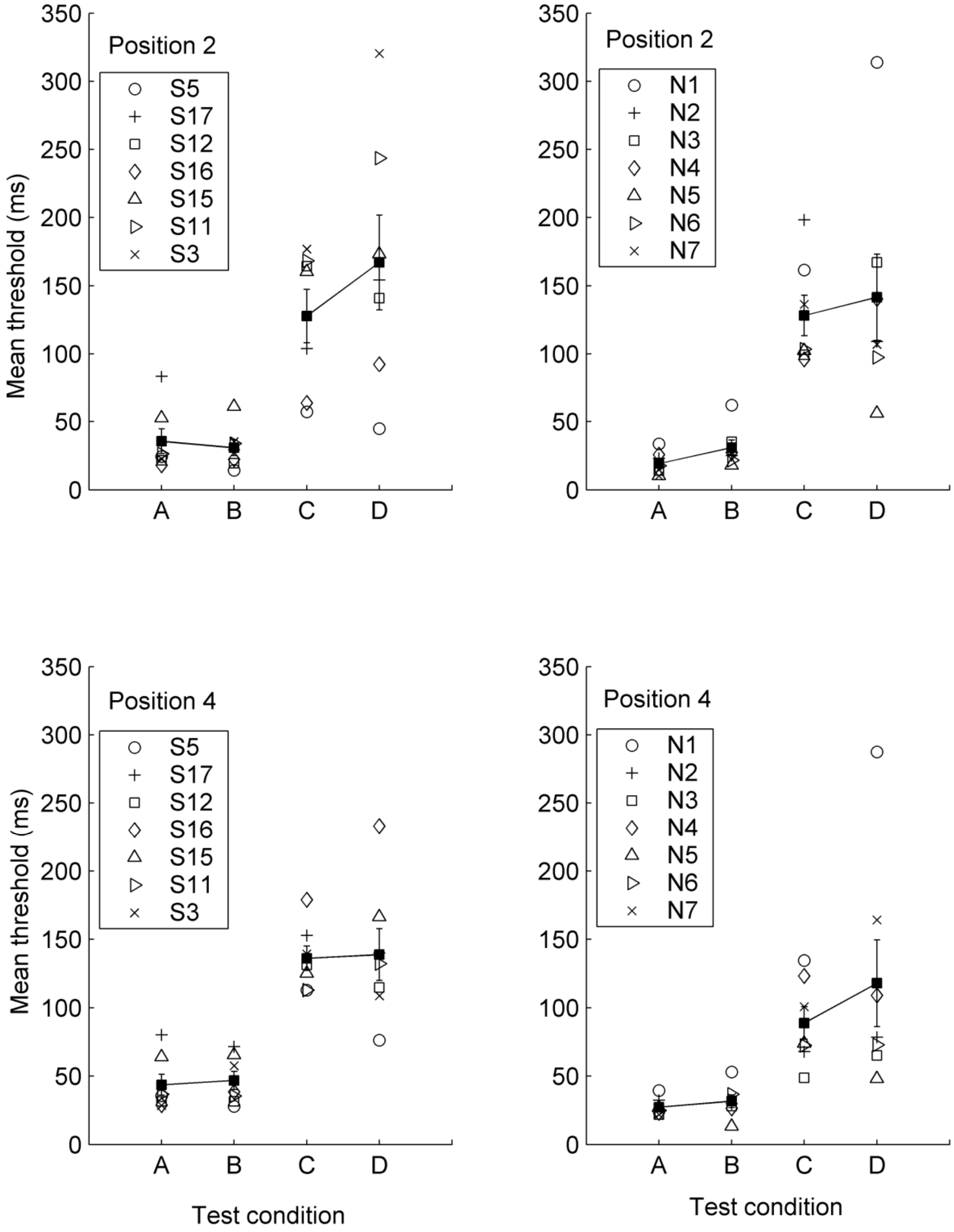

Figure 3.2: Absolute mean ITI duration discrimination thresholds in the four experimental conditions (see Figure 3.1). The top panel shows thresholds for the target interval between the second and third tones; the bottom panel shows thresholds for the target ITI between the fourth and fifth tones. Filled squares show group mean thresholds. Error bars represent standard errors.

Although anisochronous polytonic sequences appeared to present the most difficult rhythm perception condition for all listeners, discrimination thresholds were in general similar for polytonic and monotonic sequences of a specific rhythmic condition ($F(1,12) = 2.43$, $p = 0.15$). This finding indicates that rhythmically complex sequences represented more demanding listening conditions to subjects than sequences with a complex pitch pattern.

Performance trends appeared to be insensitive to the target interval's position within the tone sequence. Subjects performed similarly when the target interval was placed in position 2 or position 4, irrespective of rhythmic or pitch complexity or hearing ability ($F(1,12) = 0.52$, $p = 0.49$). This indicates that listeners were able to identify the target interval in the context of the sequence and not merely with regard to its position.

3.4 DISCUSSION

Three main findings emerged from this investigation: (i) CI and NH listeners showed similar rhythm perception ability for tone sequences as used here, (ii) pitch complexity did not adversely influence CI users' rhythm perception ability, and (iii) rhythm perception ability appeared to be insensitive to the relative position of the target interval within the tone sequence. Although the observation that CI listeners experience similar rhythm perception ability as NH listeners supports the general trend regarding CI listeners' rhythm perception abilities (Gfeller & Lansing 1991; Zeng 2004; Cooper et al. 2008; Looi et al. 2008) the investigation extends contemporary understanding by commenting on the influence of covarying pitch patterns on rhythm perception ability and CI users' ability to identify rhythmic irregularity in the context of a tone sequence.

CI users' rhythm perception ability as observed here may be associated with generally acceptable speech perception abilities afforded by modern implant devices (Wilson & Dorman 2007). Language processing has previously been shown to be subserved by a predominantly left hemisphere-based cortical network (Zatorre & Belin 2001; Zatorre et al. 2002) and in a recent neuroimaging study, where CI users performed similarly to NH listeners on a rhythm task, speech perception was also associated with predominantly left hemisphere cortical activation patterns (Limb et al. 2010). Considering that left-hemisphere-dominant activation, specifically in speech/language processing areas, has previously been observed during perception of rhythm patterns (Platel et al. 1997) the authors suggested that the intact rhythm perception ability of CI users may be related to speech perception abilities. It is thus plausible that regular activation brought on by speech stimuli may help to familiarise left-hemisphere-based cortical areas involved in rhythm perception with the nature of input derived from electrical stimulation at the auditory periphery.

The finding that rhythmic rather than pitch complexity adversely influenced the rhythm perception ability of CI users was somewhat surprising, especially considering a recent finding that amusic listeners' compromised rhythm perception abilities for pitch-varying tone sequences may be attributed to earlier pitch processing deficits (Foxton et al. 2006). Although pitch and rhythm patterns are thought to be processed separately during early analysis, the information probably combines at a later processing stage to form a unified percept (Peretz & Kolinsky 1993; Peretz & Coltheart 2003). Conceivably, the position of the deficit in the auditory pathway may influence the behavioural outcome. Degraded auditory input originates early during auditory processing for cochlear implantees, whereas for amusic listeners, the information processing difficulty appears to stem from a cortical deficit (Peretz et al. 2009). It is thus plausible that the nature of the input received by higher-order processing stages determines the neural activation patterns. Limb et al. (2010) indeed recently found more intense neural activation patterns across a more distributed cortical network in implant listeners compared to NH listeners, despite similar behavioural outcome on rhythm tasks. Heightened neural activity, especially in frontal areas, possibly reflects effortful processing and associated neural plasticity to allow successful behavioural response to degraded information. Although the findings by Limb et al. (2010) do not comment on neural activation patterns in amusic listeners, they do support the notion that degraded auditory information can elicit differential neural activity among listener groups, depending on the specific auditory deficit, which may explain the differential effect of additional pitch complexity on rhythm perception ability in amusic and CI listeners.

Lastly, the finding that rhythm perception ability was insensitive to the position of the target interval in the sequence suggests that implant listeners were able to extract rhythm cues within the context of a tone sequence, rather than focusing only on a specific intertone interval. Since rhythm can be regarded as patterns of temporal information (Samson et al. 2001), it follows that successful rhythm perception requires temporal relationships between successive tonal events to be established (Krumhansl 2000). In terms of pattern perception, rhythm perception possibly represents an intermediate processing stage – later than mere feature extraction but before semantic processing (Griffiths 2003; Koelsch & Siebel 2005). It thus appears that good early temporal resolution afforded by the implant device, as reflected by small within-channel gap detection thresholds (Moore & Glasberg 1988; Shannon 1989; Grose & Buss 2007), may propagate sufficiently to a subsequent processing level to support rhythm perception in the context of a tone sequence.

3.5 CONCLUSION

Several earlier studies have shown that CI listeners exhibit rhythm perception abilities comparable to those of NH listeners when confronted with isolated rhythmic components. However, to extend the understanding of cochlear implantees' music perception ability, rhythm perception needs to be investigated in a context that more closely represents real-world listening conditions. The finding reported here showed that CI users were indeed able to perceive rhythmic patterns equally well as the NH control group in such listening conditions. The results thus confirm that rhythm perception is not the main contributor to unsatisfactory perception in CI-mediated music listening.