Chapter 4

PERCEPTION OF SOME MELODIC CHARACTERISTICS BY COCHLEAR IMPLANT USERS

Pretorius, L.L. & Hanekom, J.J. (2011) Perception of some melodic characteristics by cochlear implant users: an exploratory study (submitted for review to *Journal of Speech, Hearing and Language Research)*

4.1 INTRODUCTION

Multichannel cochlear implants (CIs) afford many profoundly deaf listeners partially restored hearing ability. Despite deriving good speech understanding (Gifford et al. 2008; Wilson & Dorman 2008), CI users generally experience only limited music perception ability and enjoyment (Gfeller et al. 2000; Leal et al. 2003; Gfeller et al. 2003; Mirza et al. 2003; McDermott 2004; Zeng 2004; Gfeller et al. 2005). Unsatisfactory music perception may be attributed to the limited ability of the implant device to convey enough high-fidelity information relating to both the pitch and temporal domains to allow a unified and coherent musical percept to be formed (Limb et al. 2010).

Despite several studies assessing CI listeners' music perception behaviour (e.g. Fujita & Ito 1999; Gfeller et al. 2002; Leal et al. 2003; Gfeller et al. 2003; Mirza et al. 2003; Gfeller et al. 2005; Looi et al. 2008; Vongpaisal et al. 2009; Singh et al. 2009), little is known about the underlying constraints imposed by cognitive and perceptual factors, and especially their interaction, in a music listening setting. The difficulty with objectively assessing music listening success stems from the nature of musical sounds. In its simplest form music can be defined as a succession of tonal events bound together over time to form a coherent perceptual entity (Patel 2003; Limb 2006b). This implies that successful music listening involves not only analysing pitch-related and temporal information separately, but also processing their interactions.

In Chapters 2 and 3 pitch- and rhythm perception abilities of CI users were investigated at a processing level where the two dimensions are regarded to be analysed by independent processing modules within the bigger processing scheme proposed for music perception (also see Figure 1.1, Chapter 1). Since coherent music perception is the outcome of a hierarchical processing system (Peretz & Coltheart 2003), low-level deficits will undoubtedly propagate to later processing stages where processing is, in turn, linked to and influenced by other processing modules (Peretz & Coltheart 2003; Koelsch & Siebel 2005). A number of previous psychophysical assessments of CI-mediated music perception ability have investigated the perception of separate music-relevant input parameters. These studies showed that perception of pitch-related cues is severely hampered in CI users when listening in sound-field conditions (Gfeller & Lansing 1991; Gfeller et al. 1997; Kong et al. 2004; Gfeller et al. 2007; Galvin et al. 2007), while perception of temporal information (rhythm, metre and tempo) appears, if not always comparable to normal-hearing (NH) listeners' ability, at least superior to CI users' pitch perception ability (Gfeller & Lansing 1991; Gfeller et al. 1997; Kong et al. 2004; Looi et al. 2008). Given the relative importance of pitch-related information for music perception (Zatorre 2001; Zatorre et al. 2002) in especially the Western tonal tradition, it is understandable that attempts at improving the music perception ability of CI users would focus on ways to convey more intact pitch information to the electrically stimulated auditory system (Laneau et al. 2006; Milczynski et al. 2009). However, reports of slightly improved music perception when rhythm cues are available (Gfeller et al. 2002; Kong et al. 2004; Galvin et al. 2007) suggested that investigating CI perception of Western tonal music with stimuli that incorporate information from both pitch and temporal dimensions may contribute to the understanding of constraints involved in contemporary CI music perception. Studies using real-world music tokens were subsequently performed to assess the music perception ability of CI users when subjected to covarying pitch and rhythm cues (Gfeller et al. 2005; Gfeller et al. 2007).

Although such studies provided valuable empirical support for CI users experiencing only limited success when confronted with real-world music listening situations, there are some weaknesses associated with approaches that are based on recognition of familiar melodies from a specified set to evaluate perception of music-relevant information. It is not certain that the musical character of a tone sequence is retained after acoustic–electrical conversion of auditory input. Reports of reasonable individual success in studies using familiar melodies to gauge music perception ability (e.g. Gfeller et al. 2000; Kong et al. 2004; Galvin et al. 2007; Singh et al. 2009) do indeed provide useful assessment of a CI listener’s ability to match a melody to one of a limited set of stored representations supported by the available cues, but do not comment on the listener’s ability to use the available cues to drive perception of musical character and then identify a melody as one from any number of stored representations as a result. Such tasks thus do not provide a true handle for probing perception of musical character independent of memory representations, and as such offer limited possibility for tracking improvements in processing strategies or personal improvement afforded by alternative map parameters or implementation of user-specific device settings.

With many CI users expressing the wish for better music perception ability, standardised and clinically practical music perception tools are needed to allow for standardised assessment and subsequent comparison across different CI user groups, processing strategies and rehabilitation facilities. Two approaches deserve mention. Kang et al. (2009) recently put forward the University of Washington Clinical Assessment of Music Perception test as a possible clinically relevant music perception test. The test comprises pitch direction discrimination as well as recognition of commonly heard familiar melodies and identification of musical instruments from closed sets. Although this test is regarded as the most useful clinically applicable yardstick of CI music perception at present, owing to it being selfadministered and rapidly completed, it does not allow the quality of music-relevant information as conveyed by the implant to be gauged.

Cooper et al. (2008), in turn, employed the Montreal Battery for Evaluation of Amusia (MBEA) (Peretz et al. 2003) to probe CI-mediated music perception ability. This measure, based on neurocognitive principles underlying music perception, uses six tests to assess different aspects of music perception along the melodic (pitch interval, scale and contour) and temporal dimensions (rhythm and metre). It also includes a melodic memory test to assess music perception ability when confronted with co-varying pitch and rhythm information. It requires short-term memory for melodies presented in the context of the test battery, rather than the listener having to match a present melody to a representation of a familiar melody stored before becoming deaf.

It is proposed here that a music perception assessment tool that takes into account the neurocognitive principles underlying music perception, such as the MBEA test battery, is well suited to provide an overview of CI users' general music perception ability, independent of memory representations. Moreover, to assess the quality of transfer of musical character after acoustic–electrical signal conversion such tests can be expanded by including tasks that require listeners to judge syntactic congruency of short melodies. In Western tonal music, perception of musical syntactic congruency is based on the recognition of tonal relationships between successive notes. Such ability has been shown to exist in musically untrained listeners and children (Koelsch et al. 2000; Koelsch et al. 2005) and is believed to be based on inherent musical knowledge, which can be expressed without training (Trainor & Trehub 1994; Tillman et al. 2000). Tonal relationships reinforce the implied tonality and associated tonal hierarchy (Krumhansl 1979; 1990), which in turn facilitates expectancy for specific tonal events (Marmel et al. 2008).

Preceding context of a melodic line can generate expectancies for a specific musical event both at the end of a phrase and within the tone sequence itself. Both are based on the implicit knowledge of melodic key derived from the relationships between preceding notes. Out-of-key notes within a melodic line violate tonal expectancy (Brattico et al. 2006) and it follows that if such out-of-key notes can be identified correctly, the preceding notes must have built up a musical context that allows these notes to be identified as odd. Similarly, expectancy for melodic endings is generated by both the tonal and the temporal structure of the preceding context (Boltz 1989; 1993). Tonal expectancy stems from priority assigned to specific notes within a tonal hierarchy, which in turn determines their stability in a sequence (Krumhansl 1979; 1990). It has been shown that listeners expect unstable notes to resolve to stable ones to create a feeling of completion at the end of a phrase or melodic line and conversely that sequences ending on unstable (and hence unexpected) notes are judged to have a low degree of completion (Meyer 1956; Boltz 1989; Bigand & Pineau 1997). Temporal relationships between successive notes have been shown to generate similar expectancies within the temporal domain (Boltz 1993; Nittono et al. 2000) and contribute to judgements of melodic completion possibly through creating accents that direct a listener’s attention to a specific event (Boltz 1989; Jones 1993; Boltz 1993). Melodies ending on prolonged final notes, especially when these are stable notes, have been shown to be judged more complete than those without a lengthened final note (for a description see Boltz 1989).

Melodic expectancies contribute to the build-up of melodic context for music listening, which can facilitate processing of future bottom-up information (Schulkind 2004). It thus follows that without proper perception of the underlying musical information a musical context will not be established, which in turn may lead to impaired higher order auditory processing of music. Since CI listeners experience only limited melodic perception, probing their perception of melodic expectancy as a marker of melodic context synthesis may be a valuable approach.

According to Koelsch et al. (2004) syntactic violations elicit similar brain responses in CI listeners and NH control subjects, indicating that the neural mechanism underlying detection of music syntactic congruency is present in CI listeners. The finding suggests that tasks which require perceptual judgements to be made based on musical expectancy may (i) provide a useful indication of CI listeners’ ability to extract musical context from a sequence of preceding events when presented with unfamiliar real-world-like melodies and (ii) contribute to the development of new or further refining of existing clinically practical test batteries to compare music perception performance across CI user groups and gauge perceptual outcomes of future processing strategies.

The investigation described in this chapter explores CI users' ability to perceive simple, coherent melodies in which pitch and rhythm information contribute simultaneously to form a unified musical token. It is based on the notion of the music processing system being a modular, hierarchically organised perceptual system as a departure point. According to this view, forming a melodic percept is the outcome of a processing step that occurs beyond separate pitch and rhythm analysis but needs direct input from these processing modules. Given the multifaceted nature of music as well as the human auditory processing system's ability to adapt to less-than-optimal input, it is worthwhile to explore CI users' ability to judge musical character of simple melodies as a window into the system-dependent processing of electrically mediated auditory input. In view of this explorative approach, CI listeners' judgement of musical syntactic congruency either at the end of melodic lines (Experiment 1) or within the tone sequence (Experiment 2) is assessed. The two perceptual tasks presented in the respective experiments described here are offered as alternative approaches to assessing the extent to which useful musical context is built up in CI-mediated sound-field hearing.

4.2 EXPERIMENT 1: PERCEPTION OF MELODIC COMPLETION

In Western tonal music, relationships between seven diatonic tones are used perceptually to establish the key of an unfolding melody. Within the framework of the established key each tone is perceptually assigned a priority, which translates to a certain level of melodic stability within the key (Krumhansl 1979; 1990). The relative stability or instability of a note within a key, coupled with rhythmic patterns, in turn generates expectancy in the listener for a specific note to follow, based on the preceding context.

Expectancy judgements have previously been used to investigate the information processing and neurocognitive principles underlying music perception (e.g. Boltz 1989; Boltz 1993; Bigand & Pineau 1997; Koelsch et al. 2000; Koelsch et al. 2005; Brattico et al. 2006; Marmel et al. 2008). Brattico et al. (2006) showed that judgements regarding congruency violations at the end of melodic phrases are possibly facilitated by the output of earlier feature extraction, which suggests that perception of harmonic belonging may depend on prior successful perception of musical scale belonging. In the context of CI music perception research, markers of processing stages can be usefully applied to investigate to what extent the information conveyed by the implant device supports music perception. Experiment 1 investigated CI users' perception of melodic completion to determine whether the information conveyed by the implant device allows successful processing at this relatively advanced processing stage.

Since CI users find melody recognition easier when both pitch and rhythmic cues are available (Gfeller et al. 2002; Kong et al. 2004; Galvin et al. 2007), it was deemed valuable to investigate whether rhythmic information can influence perception of melodic completion to a similar extent as for NH listeners (Boltz 1989; 1993). Perception of melodic completion was therefore investigated with and without temporal cues to signal melodic completion.

4.2.1 Methods

Subjects

The CI test group consisted of seven post-lingually deafened adult users (mean age = 49.3 yrs) of the Nucleus 24 or Nucleus 22 electrode array device. All participants had more than two years' experience with their devices and had experienced profound hearing loss for more than 10 years. CI listeners all used the Freedom processor; five fitted with the ACE processing strategy and two with the SPEAK strategy. Three of the participants were implanted bilaterally, but only the ear subjectively regarded as the better of the two was used during testing. Only one of the CI participants had formal music training prior to the onset of deafness (S10), while two others had been choir members during their youth (S3 and S8). Two more (S21 and S18) indicated that they regularly try to listen to music or the radio, despite perception often being unsatisfactory. All but one (S21) of the participants had participated in earlier CI research at our laboratory. Other relevant demographic details are provided in the Appendix (Table A1 and explanatory description). An overview of participants' general listening success with the CI, as rated by the treating audiologist, is given in Table A2. The questionnaire was adapted from the Abbreviated Profile of Hearing Aid Benefit questionnaire (Cox & Alexander 1995).

Seven age-matched NH control group listeners (mean age = 48.6 yrs) also participated in this investigation. Normal hearing was defined as achieving audiometric thresholds of 30 dB HL or better at five octave frequencies from 250 to 4000 Hz. Subject N3 achieved only a 35 dB HL threshold at 4000 Hz , but since stimulus frequencies never exceeded 1600 Hz (see section on stimuli), he was included in the control group. 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. Two of the control group listeners had previously participated in research at our laboratory.

Owing to the exploratory nature of this investigation, the same group of listeners participated in all tasks described here. All participants (CI and NH) gave written informed consent prior to commencement of the investigation according to the requirements of the relevant ethics committee. Participants were compensated for their time at the conclusion of the investigation.

Stimuli: general considerations

The tasks described here are suggested as a first step towards probing a subjective perceptual quality in as objective a manner as possible. The explorative nature of the investigation creates the backdrop for the selection of test stimuli. Firstly, aural training material (Horacek & Lefkoff 1970; University of South Africa 1970; Van Zuilenberg 1996) was considered to provide a suitable pool of short, simple, single-voice melodies with easy-to-follow pitch contours and rhythmic structure, especially since all stimuli were chosen from the primary grades' material. It was deemed more feasible to use existing melodies that were all of the same difficulty level than composing melodies anew and run the risk of them not being of similar perceptual difficulty. Secondly, use of aural training material avoids the risk of the stimuli being familiar to listeners, but simultaneously provides the assurance that the tokens are true melodies created according to established principles of melody composition. All stimuli were reviewed for their melodic quality and 'listening ease' by a colleague at our university's music department during design of the investigation and comments and recommendations were duly addressed. Thirdly, only melodies in simple time, starting on the tonic and of which the shortest note value was an eighth note were included in the stimuli set. This was done to allow listeners a reasonable amount of melodic information on which to base their decisions, without burdening them with too rapid or intricate rhythmical patterns. The resulting set of 20 melodies was regarded to provide a suitably varied stimuli set without dragging on too long and so risk listeners losing concentration, based on personal experience working with the listener profile available at our implant centre over several years. All stimuli and tasks were piloted with a NH listener before commencement of the study.

Stimuli: specific considerations

A core set of 20 melodies were adapted from material used for graded aural training in music education (as described above) to ensure novelty to listeners. The melody set is shown in Figure 4.1(a) and (b).

Although it is customary to perform experiments of melodic completion perception with chord sequences, unaccompanied tonal melodies consisting of pure sine tones generated in Matlab 6.5 were used in this experiment to prevent unnecessary spectral information from introducing additional processing difficulty. Melodies were between 15 and 23 notes long (average length = 17.4 notes) and contained only half, quarter and eighth notes, separated by 90 ms gaps (Cowan 1984). Amplitude ramps (30 ms) were included at the beginning and end of each tone to reduce onset clicks. Melodies were presented in simple duple, triple or quadruple time at a tempo of 150 beats per minute.
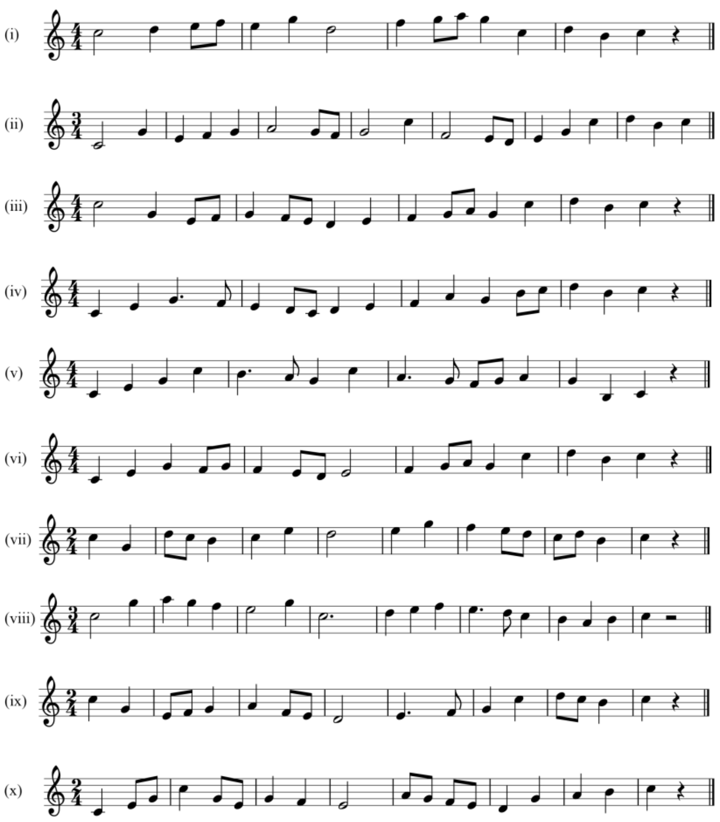
Figure 4.1a: Ten melodies with leading tone–tonic final progressions, from the set of 20 melodies used in the melodic completion task.

Melodies always started on the tonic, but never exceeded a 1.5-octave span. Since each melody was presented in C, F and G major, the frequency span ranged from 220 Hz (A3) to 1318.51 Hz (E6). This range falls roughly within the cochlear region where frequency filters of the speech processor are linearly spaced. An arpeggio, consisting of the tonic, mediant, dominant and end-octave tonic tone, was presented before each melody to establish its implied key.
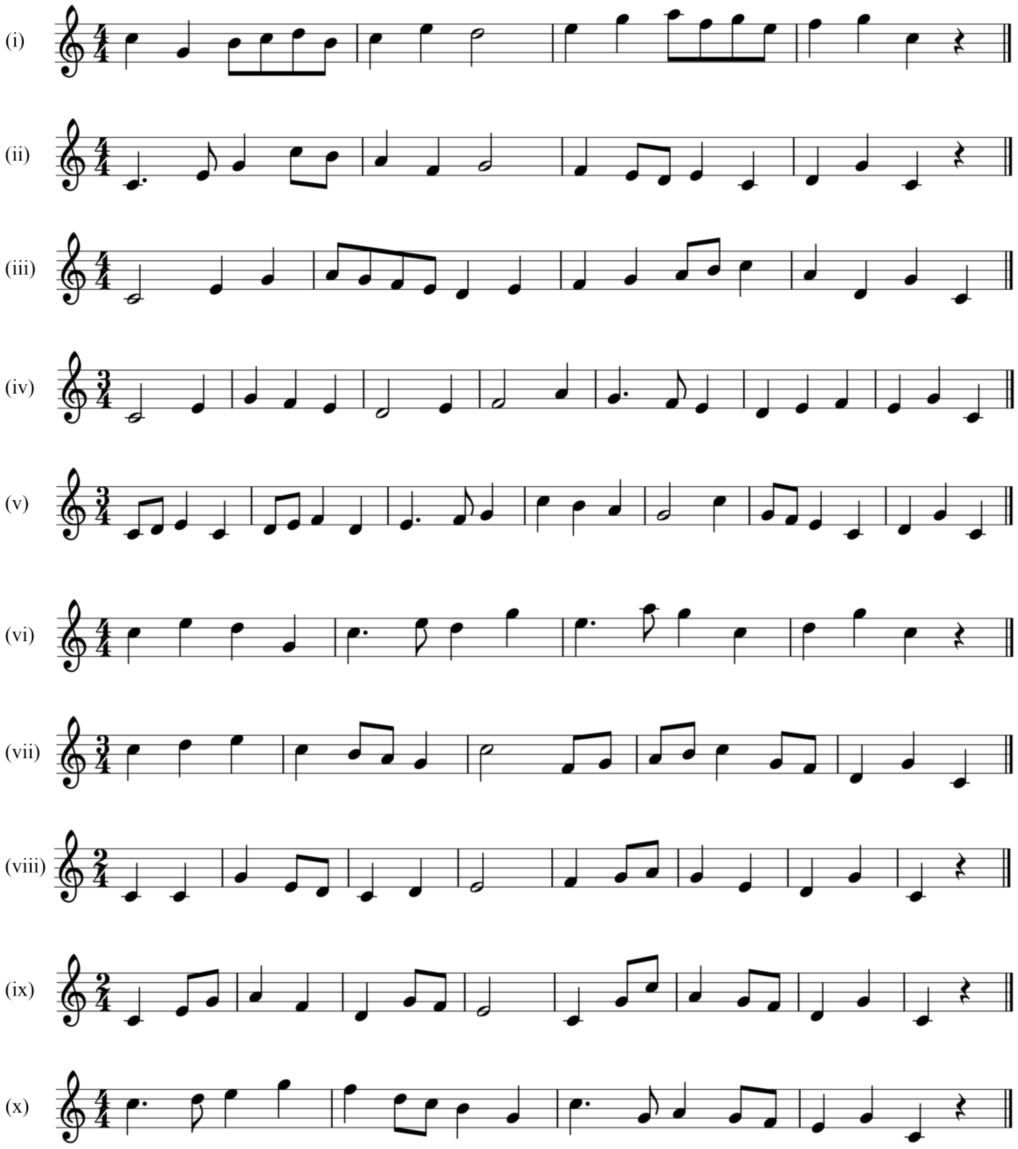
Figure 4.1b: Ten melodies with dominant–tonic final progressions, from the set of 20 used in the melodic completion task.

Each of the 20 melodies was presented with a complete and incomplete ending during the course of the experiment. However, each of these (40) melodies was presented in three keys and the stimuli sets were therefore divided into subsets of ten melodies each so as to prevent an experiment run from becoming too tedious. Complete endings were created by using either a leading tone–tonic or a dominant–tonic progression, while the order of the last two notes was reversed to create incomplete endings. To prevent progression direction from unduly being used as a completion cue, owing to melodies with complete and incomplete endings differing only with regard to the order of their last two notes, equal numbers of samples of both progression types (leading tone–tonic and dominant–tonic) were presented in each subset of complete melodies and then paired with a complementary subset of incomplete melodies. Examples are shown in Figure 4.2(a) to (d).
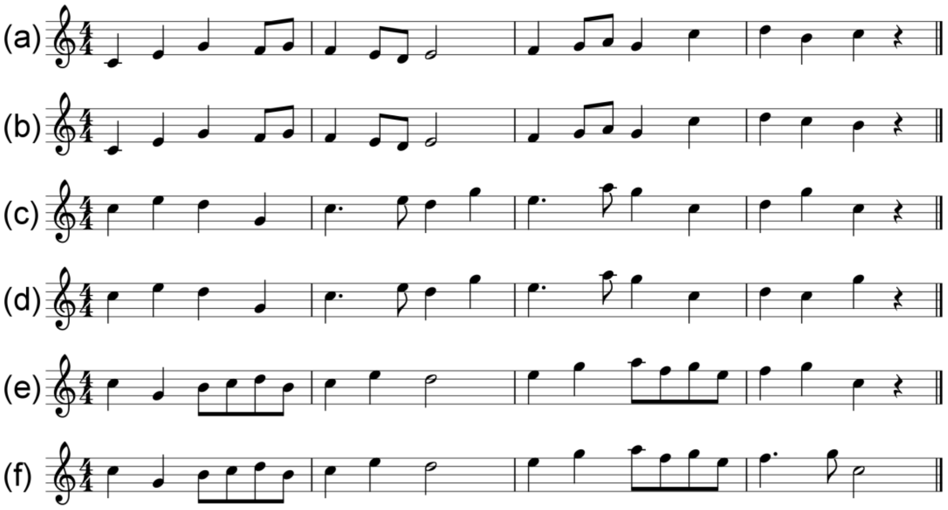
Figure 4.2: Examples of stimuli used in the melodic completion task. Melodies with leading tone–tonic and dominant–tonic final progressions are shown in staves (a) and (c) respectively, while their complements with incomplete final progressions are shown in (b) and (d). Staves (e) and (f) show an example of a melody without and with a lengthened final note.

Perception of melodic completion was investigated both with and without temporal cues in the final progression to determine whether rhythmic information would influence completion judgements to the same extent as for NH listeners. However, separate stimuli sets (consisting of the core set of 20 melodies) were compiled for testing with and without rhythmic cues and were not mixed during an experiment run. For stimuli without rhythmic cues, the notes of the final progression were equally long (quarter notes), while for stimuli with rhythmic cues the length of the last note (half note) was four times that of the penultimate one (eighth note). Examples are shown in Figure 4.2(e) and (f).

Since all experiments were to be conducted in sound field, with subjects using their clinically assigned speech processors and settings, it was important to present stimuli at comfortably audible loudness levels, rather than at a fixed intensity. Participants therefore had to perform a loudness estimation task during which the subjective loudness of a 1 kHz tone presented at 10 intensities between minimum and maximum sound-field loudness had to be estimated. Each intensity level was presented 20 times in random order and participants had to assign a value of between 1 and 100 to describe the tone’s loudness. A value of 1 corresponded to a sound that was just audible, while 100 corresponded to one that would be regarded as being of maximum comfortable sound-field loudness. A comfortably audible loudness level was defined as between 50% and 70% as determined from a resulting loudness estimation curve.

Loudness balancing was performed in order to rule out any confounding loudness cues that could influence melody perception performance. Frequencies selected at 100 Hz increments from the range relevant to the melody perception tasks were loudness balanced to the 1 kHz tone at the level earlier determined to represent a comfortably audible loudness. Participants were required to adjust each of the probe tones to sound equally loud to, just louder and just softer than the 1 kHz reference tone. Probe frequencies were presented in random order in triplicate during each adjustment task and the average deviation from the reference intensity was subsequently calculated from the nine loudness judgments. Presentation intensities for each frequency were stored in a look-up table for use during the melody perception tasks.

Experimental procedure

To complete a full set of data responses for a specific experiment condition, participants were required to complete two experiment runs of three task repetitions each. During a single experiment run each task consisted of ten melodies with complete endings, paired to a complementary set of ten melodies with incomplete endings and presented in three keys each, for a total of 60 trials per task. The remaining ten melodies with complete endings from the core set of 20 were presented similarly during the second experiment run and again paired to a set of ten complementary melodies with incomplete endings. A full data set thus consisted of responses to 180 melodies with complete endings (20 melodies x 3 keys x 3 repetitions) and an equal number of melodies with incomplete endings.

All experiments were performed in sound field. Stimuli were presented through a Yamaha MS101 II loudspeaker, which was placed approximately 1 m in front of the listener on the side of the test ear. Subjects were instructed that they would hear a short melody and had to judge whether its ending sounded complete or incomplete. An incomplete ending was described as one that would end abruptly or unexpectedly, leave the listener “hanging”, or induce a feeling in the listener of wanting to add additional notes to round off the melody. Conversely, a complete ending was described as one that would leave the listener satisfied that nothing more was needed or expected to round off the melody to a meaningful unit.

The experiments were of a yes/no design. Participants were required to indicate whether the ending of the melody sounded complete by selecting either the “yes” or “no” button on a graphical user interface (GUI). All stimuli were presented in random order. Three practice trials were presented prior to commencement of each task. All practice trials had to be completed successfully before data recording would start. Participants were aware of the practice trials, but did not know how many would be presented before each task. No feedback was provided, but participants could monitor the task progress via a progress bar included on the screen. All tasks were self-paced and a new melody would not be presented before listener response had been received to the previously presented stimulus. Each task repetition lasted approximately 20 minutes and a single experiment run thus lasted almost an hour.

Although these were quite long sessions, participants were urged to take short breaks between each of the tasks.

4.2.2 Results

Figure 4.3 shows average completion perception ability of the respective listener groups (NH and CI listeners) for each combination of melodic ending and progression type (experiment condition A–D). Data responses were organised according to experiment condition and each listener’s performance was thus calculated as the percentage correct responses over 90 trials in each of the four conditions (10 melodies x 3 keys x 3 repetitions x 4 conditions).[6] The top and bottom panels represent results for melodies with and without rhythmic cues respectively. Melodic ending (complete or incomplete), progression type (leading tone–tonic or dominant–tonic pairs) and final note duration (lengthened or not lengthened) were defined as within-subject variables, while hearing ability (CI or NH) was defined as the between-subjects variable in a three-way repeated measures analysis of variance.

NH listeners achieved (average) scores of 83–96% correct, irrespective of final note duration, while CI listeners performed close to or below chance level for all but dominant–tonic progressions in both note length conditions. The difference in performance level between listener groups was highly significant ($F(1,12) = 104.70$, $p < 0.001$).

Although no significant main within-subjects effects were found in either of the listener groups, several interactions were found to be significant at the 0.05 decision level. Significant interaction between hearing ability and progression type ($F(1,12) = 6.57$, $p < 0.05$) indicates that performance was differently influenced by the progression type for the two listener groups.

---

[6] S8 completed only 45 trials (five melodies x three keys x three repetitions) in each of the four conditions owing to his limited time availability.

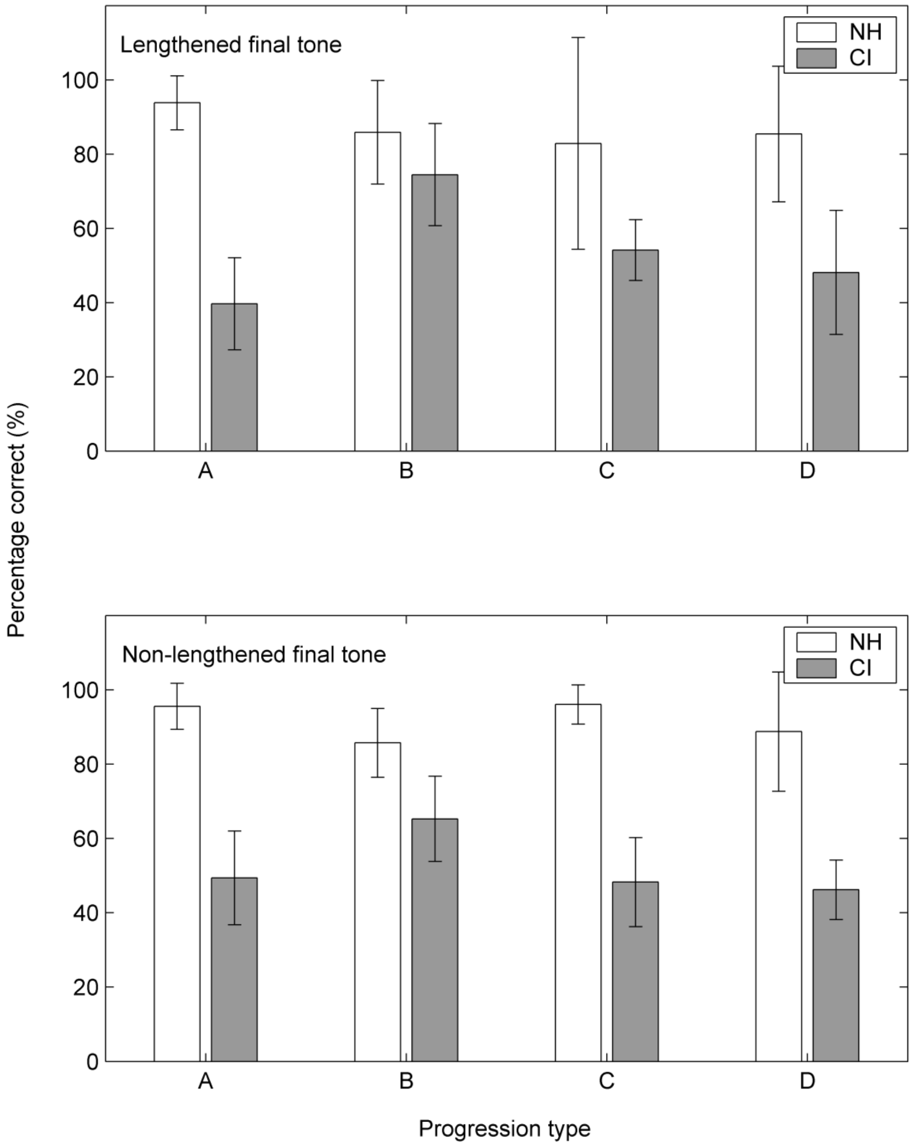

Figure 4.3: Average performance across listener groups in melodic completion task, with final note lengthened and not lengthened in top and bottom panels respectively. A = leading tone–tonic; B = dominant–tonic; C = tonic–leading tone; D = tonic–dominant. Error bars indicate one standard deviation from the mean.

CI listeners were better able to use dominant–tonic progressions for judgements regarding melodic completion, while NH listeners used leading tone–tonic progressions better to this end.

The significant interaction between hearing ability, melodic ending and progression type (F(1,12) = 10.94, $p < 0.01$) shows that, as for the interaction between hearing ability and progression type, progression type influenced performance in the two listener groups differently. Moreover, progression type exerted its most pronounced influence in melodies with complete endings. Taken together, these results indicate that CI listeners were better able to recognise complete endings involving dominant–tonic progressions, while NH listeners were better able to recognise complete endings that involved leading tone–tonic progressions. A similar effect was not seen for melodies with incomplete endings.

Significant interaction was also found for final note duration and progression type (F(1,12) = 7.19, $p < 0.05$). Both listener groups found judging melodic ending slightly easier when a lengthened final note was used together with a dominant–tonic progression than with a leading tone–tonic progression. However, a similar facilitating effect of progression type was not seen for melodies without a lengthened final note.

4.2.3 Discussion

Results of Experiment 1 show that information required for successful and reliable judgement of congruent melodic completion was not available to CI listeners to the same extent as to NH listeners. The effect of hearing ability appears to have masked any within-subject variables from significantly influencing perceptual outcome during these tasks, which may point to either insufficient or inappropriate information reaching higher-order cortical processing after electrical stimulation of the auditory system. Judgement of the congruency of a melodic ending depends largely on the assignment of tonal priorities according to an implicit tonal hierarchy, which in turn is based on successful perception of pitch relationships. Considering the limited pitch resolution afforded by the CI device and ensuing broad neural activation patterns, such insufficient or inappropriate information reaching an advanced music processing stage may result from impaired feature extraction already during early auditory processing stages, especially within the framework of a hierarchical, modular music processing system (Zatorre et al. 2002; Peretz & Coltheart 2003; Koelsch & Siebel 2005).

The observation of CI users' better performance with regard to melodies with dominant–tonic final progressions compared with those with leading tone–tonic progressions may be attributed to the wider frequency difference between the two consecutive final tones. A dominant–tonic progression is seven semitones apart, while a leading tone–tonic progression spans only two semitones. Depending on the key (which would dictate the frequency of the tonic) two tones seven semitones apart would amount to a 130–195 Hz frequency difference – more than one filter width of the speech processor – whereas a leading tone–tonic progression would not exceed 47 Hz. Previous studies regarding musical pitch perception have found that, on average, CI users struggle to reliably discriminate tonal intervals of less than five semitones (Gfeller et al. 2007; Galvin et al. 2007; Looi et al. 2008; Pretorius & Hanekom 2008), which has been suggested to be related to electrode-associated filter width and resulting overlapping neural activation patterns (Pretorius & Hanekom, 2008). However, the wider frequency separation resulted in a significant effect only in melodies with complete endings. If preceding melodic context were built up accurately, the same effect would have been expected for melodies with incomplete endings and it is hence reasonable to infer that the effect of wider frequency separation seen with complete endings may have resulted more out of bias towards positive responses than proper interpretation of preceding melodic context. This underlines the uncertainty experienced by CI users when having to make melodic judgements during music listening.

4.3 EXPERIMENT 2: PERCEPTION OF MUSICAL KEY VIOLATION

Experiment 1 showed that sufficient information for successful judgement of melodic completion is not available to CI users. Use of contextual melodic information was therefore investigated also at a different music processing stage. Experiment 2 sought to determine whether CI users are able to establish pitch relations between tones of a melody such that violation of the underlying musical key can be inferred. The perception of tones belonging to a specific musical key is governed by implicit rules of tonality of Western tonal music (Krumhansl 1990) and has been shown to be processed pre-attentively by listeners without music training (Tillman et al. 2000; Brattico et al. 2006).

Perception of musical key belonging represents an intermediate stage of music processing, occurring later than feature extraction, but earlier than syntactic judgement (Koelsch & Siebel 2005). Since pitch information thus needs to be integrated according to the pattern established by preceding tones, (global) contour analysis provides a reference point for (local) interval perception (Peretz 1990; Stewart et al. 2008). Considering the hierarchical framework proposed for music processing (Peretz & Coltheart 2003; Koelsch & Siebel 2005; Brattico et al. 2006), it follows that melodic character would not be perceived without successful earlier (primary) feature extraction. When applied to CI-mediated music perception research, an approach that considers stepwise build-up of musical context may provide useful beacons in search of establishing improved understanding of the underlying mechanisms that govern CI-mediated music perception.

Perception of musical key belonging, as a measure of melodic character perception, was investigated under two experimental conditions. Since accents serve as perceptual markers within a melody (Jones 1993), key-deviant notes were placed in either an accented or an unaccented position. Perceptual accents are generated relative to the preceding melodic context, both at pitch level (contour and interval patterns) and temporal level. Considering that coinciding pitch and temporal accents increase the saliency of melodic markers (Jones 1993), CI listeners' perception of musical key violation was investigated with and without melodic accents to assess the extent to which melodic context can assist music perception by this listener group.

4.3.1 Methods

Subjects

The same group of CI and NH listeners who participated in Experiment 1 also participated in Experiment 2. Informed consent granted at the beginning of the investigation covered participation in both experiments. Subjects were again compensated for their participation.

Stimuli

The same core set of 20 melodies used for Experiment 1 was used during Experiment 2. Again all stimuli consisted of pure sine tones, generated in Matlab 6.5, at frequencies associated with tones between A3 and E6 and at intensity levels as determined for Experiment 1. All melodies were presented with final progressions that satisfied the requirements for harmonically complete endings and without any lengthening of the final note. Key deviations were created by tuning a target tone one semitone up or down, to violate belonging to the implied diatonic scale, but not that of the more generally implied chromatic scale. An arpeggio, consisting of the tonic, mediant, dominant and end-octave tonic tone, was presented before each melody to establish its implied key.

Target tones were placed well into the melody but not too close to the final progression, to ensure that sufficient melodic context would be available to direct listeners' decisions. Tonic and dominant tones (and if possible all three members of the tonic triad) featured melodically earlier than the target tone to reinforce the tonal context established by the broken chord.

Target tones at unaccented positions qualified as being neither pitch nor temporally conspicuous according to Jones's (1993) Joint Accent Structure, while accented positions satisfied the requirements for both pitch and temporal accent. The melody sets are shown in Figure 4.4(a) and (b).
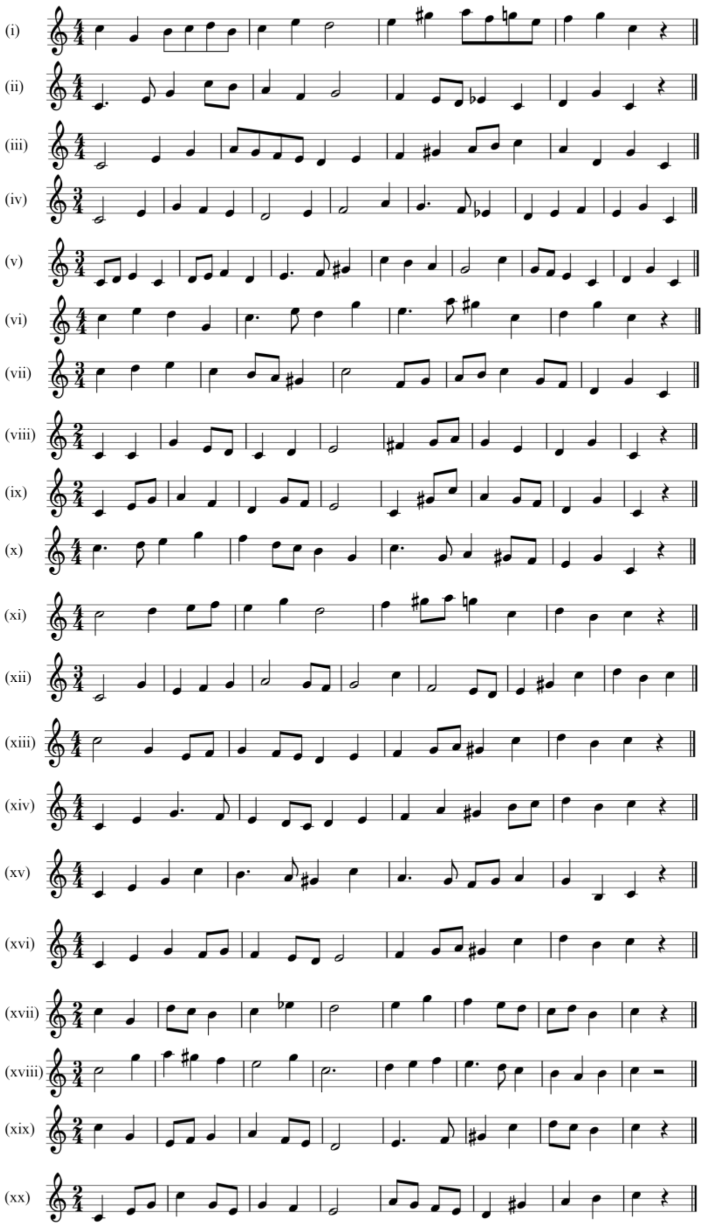
Figure 4.4a: Set of 20 token melodies with the target tone in an unaccented position.
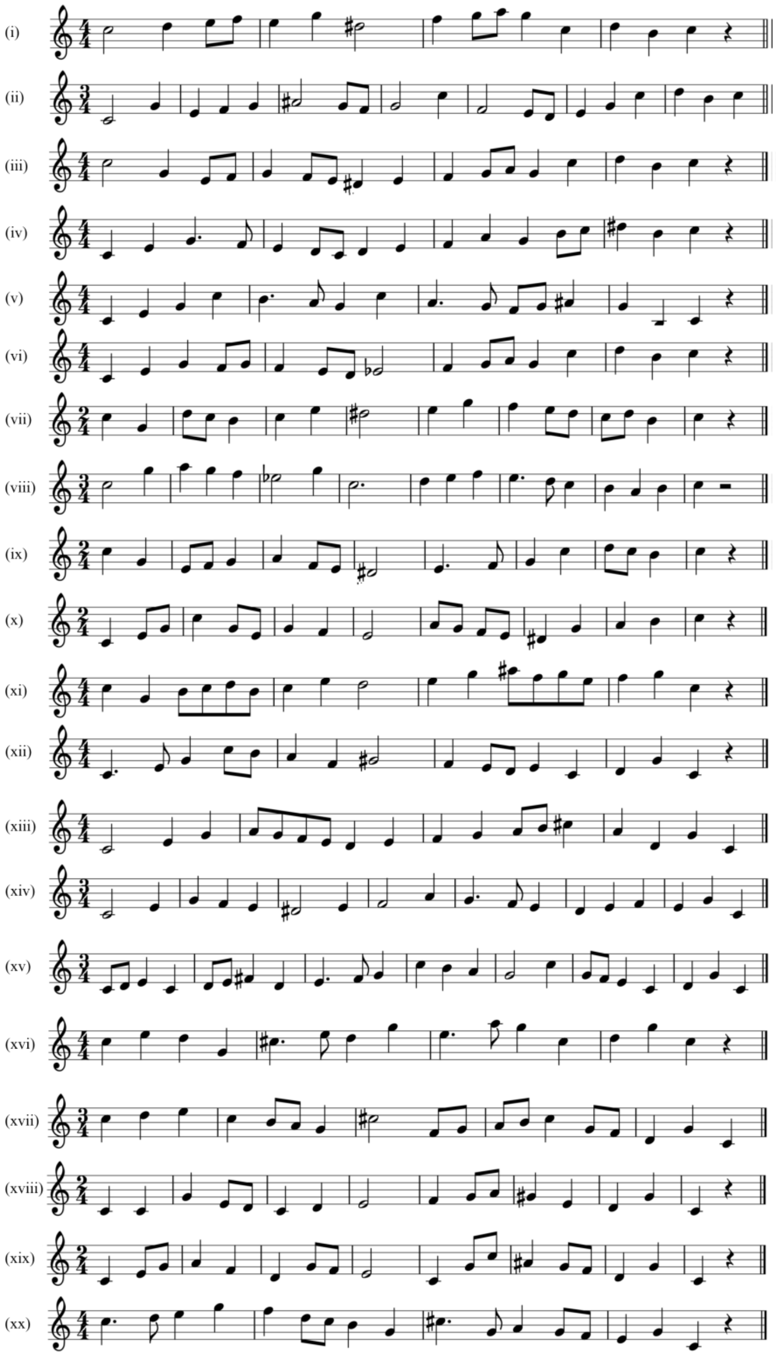
Figure 4.4b: Set of 20 token melodies with the target tone in an accented position.

Similar to the set-up used in Experiment 1, each of the 20 melodies was again presented in C, F and G major, resulting in a total stimulus set of 60 melodies. As during Experiment 1, the set was divided into two subsets, each containing an equal number of melodies ending with leading tone–tonic (upward) and dominant–tonic (downward) progressions. A melody subset containing key-deviant target tones was always paired with a complementary subset of melodies which did not contain any key deviant tones of the same accent condition (either accented or unaccented target tones), so that a single melody would not be presented with and without a key violation during the same experiment run.

Experimental procedure

The procedure for Experiment 2 was similar to that of Experiment 1. Within an experimental condition (accented or unaccented) subjects had to complete two experiment runs of three task repetitions each to complete a full set of data responses. Each task comprised 60 melodies, of which 30 (10 melodies presented in three keys each) contained key-deviant target tones. A full data set per condition thus comprised responses to 180 melodies containing a key-deviant target tone (20 melodies x 3 keys x 3 repetitions) and 180 melodies without a key-deviant target tone.

The experiment was performed in sound field in a quiet room without background noise. A Yamaha MS101 II loudspeaker was placed approximately 1 m in front of the listener on the side of the test ear. Subjects were instructed that they would hear a melody from the same set used during Experiment 1 and had to judge whether a specific tone violated the tonality rules of the underlying key. If the tone sounded out-of-key, subjects had to select the “yes” button on the GUI, while an in-key tone had to be associated with the “no” button. The position of the target tone was indicated on screen by a visual marker displayed for the duration of each melody, while a second marker indicated the progress of the melody. When the two markers aligned they turned red to alert subjects to the target tone being presented. All stimuli were presented in random order.

Practice trials, feedback and progress monitoring options were implemented as for Experiment 1. Experiment sessions again lasted approximately one hour and subjects were urged to take short attention breaks as necessary.

4.3.2 Results

Figure 4.5 shows average key violation perception scores for CI and NH listener groups. Results for melodies with and without key-deviant target tones are shown separately in both accented (top panel) and unaccented (bottom panel) conditions. Individual scores were calculated as the average percentage correct after 180 stimulus presentations (20 melodies x 3 keys x 3 task repetitions) in each experiment condition. Key violation status (in-key or out-of-key) and accent condition (accented or unaccented target tone) were defined as within-subject variables, while hearing ability (CI or NH) was defined as the between-subject variable in a two-way repeated measures analysis of variance.

Similar to their performance in Experiment 1, CI listeners performed significantly worse during the key violation perception experiment than NH listeners ($F(1,12) = 227.46$, $p < 0.001$). It is interesting to note, however, that performance trends were similar across the two listener groups. Both groups performed 12–15% worse for melodies with a key-deviant target tone present than for those without key violation. Key violation status was subsequently found to have a statistically highly significant effect on perception ability ($F(1,12) = 14.994$, $p < 0.005$). Perception ability was not significantly different for accented or unaccented target tones ($F(1,12) = 0.314$, $p > 0.05$). No interactions between variables were significant at the $p < 0.05$ decision level.

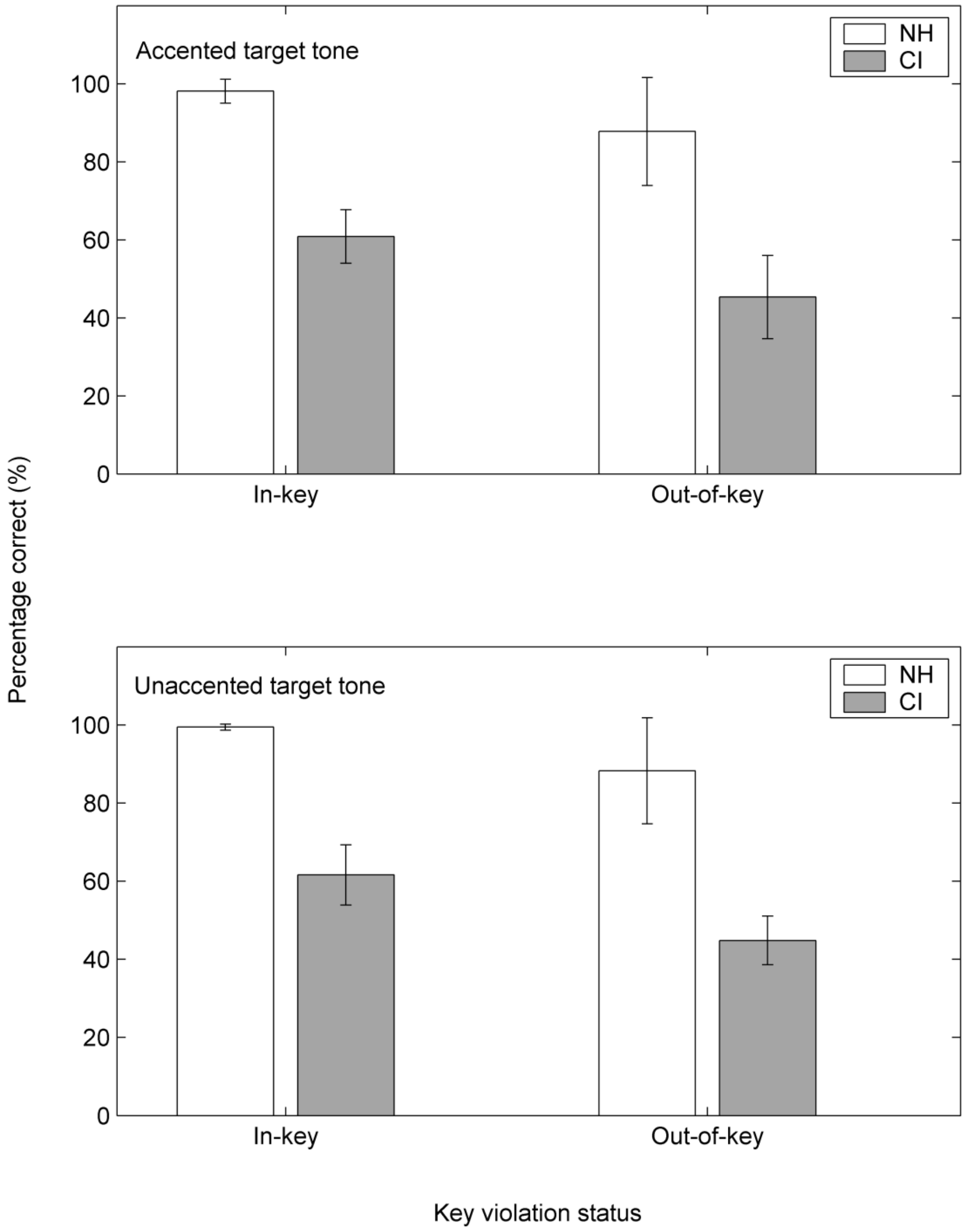

Figure 4.5: Average performance across listener groups in key violation task for accented and unaccented target tones in top and bottom panels respectively. Error bars indicate one standard deviation from the mean.

In view of these results, a brief experiment was performed using musical scales and arpeggios as stimuli. This was done to determine whether CI users are better able to use melodic context for recognition of a key-deviant tone when listening to a highly regular and expected tone sequence. Stimuli consisted of ascending and descending C, F and G major scales and arpeggios, each randomly presented with and without a key-deviant target tone. As for the previous experiment, a key violation was generated by tuning the target tone a semitone up or down from its normal pitch. Since accent condition did not significantly affect perception ability in the melody task, no such distinction was used for the present experiment. A complete stimulus set contained 24 tone sequences, of which subjects had to complete three repetitions. A full experiment run lasted approximately 20 minutes. Experiment design and set-up, task instructions, tone characteristics and frequency ranges were otherwise as for the melody task.

Results of the scale/arpeggio task are shown in Figure 4.6. Individual scores were calculated as the average of 36 data responses (12 stimuli x 3 repetitions) per condition. Each bar represents the group average per condition. Results were compared to those obtained during the melody task (accent condition), using violation status (in-key or out-of-key) and stimulus type (scale or melody) as within-subject variables and hearing ability (NH or CI) as between-subject variable during a two-way repeated measures analysis of variance. Since the target tone in the scale task was a tonic triad member (mediant or dominant), which qualifies as a pitch accent owing to tonal priority (Meyer 1956; Krumhansl 1979), results from the accented melody task were used for comparison during analysis.

Similar to the effect seen in the melody task, stimuli with a key violation resulted in significantly worse perception ability than those without ($F(1,12) = 8.72$, $p < 0.05$). The trend was similar across both listener groups, as confirmed by the non-significant interaction between hearing ability and violation status. Neither stimulus type nor its interaction with violation status was found significant at the 0.05 decision level. As in both earlier experiments, hearing ability exerted a highly significant influence on perception ability ($F(1,12) = 146.33$, $p < 0.001$).

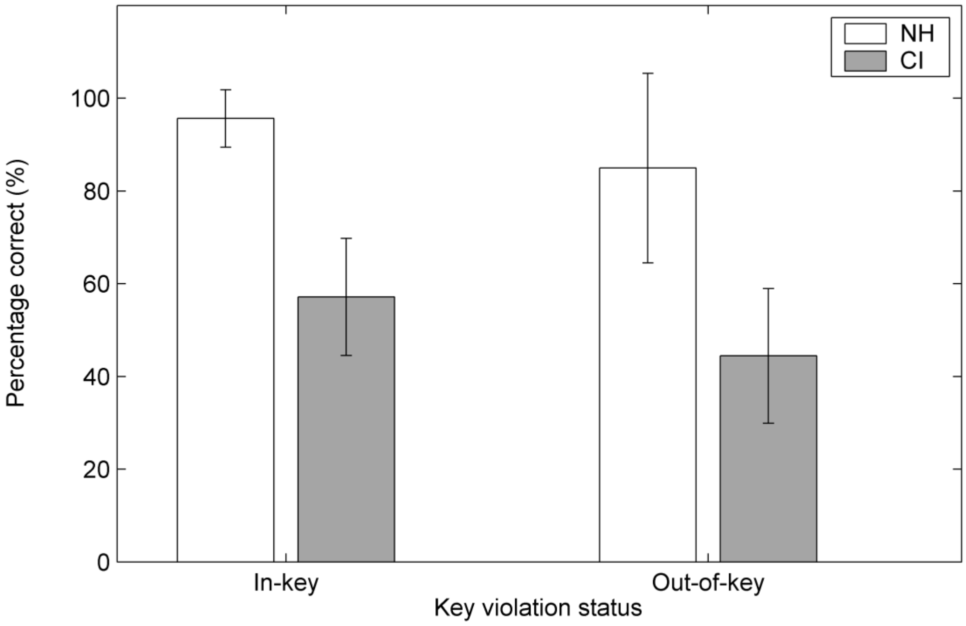

Figure 4.6: Average performance across listener groups in scale/arpeggio task. Error bars indicate one standard deviation from the mean.

4.3.3 Discussion

Taken together, results from Experiment 2 show that CI listeners could not access information needed for judgement of key violation, whether in the context of a melody or a more regular scale/arpeggio, to the same extent as NH listeners. This indicates that CI listeners were not able to build up sufficient melodic context from an unfolding tone sequence to drive melodic perception as experienced by NH listeners. Since a pitch difference of only one semitone, which corresponds to a frequency ratio of 1:1.06, had to be perceived during this experiment, it is likely that the pitch information reaching the cortical music processing system was inadequate to support perception of unfamiliar, real-world melodies. Compromised pitch quality may in turn be attributed to the limited spectral resolution available following conversion of acoustic to electrical signals already during early auditory processing (Shannon et al. 1995; Smith et al. 2002; Shannon 2005).

The finding that accent condition did not afford significantly different perception ability supports the powerful contribution of pitch information to successful music perception (Zatorre 2001; Zatorre et al. 2002). According to Jones's (1993) theory of joint accent structure, coinciding pitch and rhythmic accents serve as perceptual markers during melody perception, thereby promoting contour perception, which in turn facilitates local interval perception (Peretz 1990). However, it appears that pitch information may have masked facilitating effects of the accent structure of the tone sequence, with contrasting end results. For NH listeners, the available pitch information seem to have adequately signalled key violation status, without support from additional melodic markers needed for judgement; for CI listeners, however, the effect of compromised pitch information reaching this processing stage may have proved too strong for cues from the accent structure to render any helpful effect. This underlines the need for improved pitch resolution of the implant device.

The marked performance difference associated with violation status seen in both listener groups was somewhat surprising. If the results were due to pitch resolution alone, performance would be expected to have been similar for tone sequences either with or without a key violation. As such, NH listeners would have been expected to perform equally well – and CI listeners equally poorly – in both conditions, owing to the respective groups' earlier demonstrated frequency discrimination thresholds (for NH listeners see Wier et al. 1977; for CI listeners see e.g. Gfeller et al. 2007; Pretorius & Hanekom 2008). The finding that both listener groups performed worse on tone sequences with key-deviant tones (irrespective of stimulus type) suggests that in the presented context key-deviant target tones may not always have been unambiguously salient (Brattico et al. 2006). The result may thus also reflect response bias introduced by salience ambiguity, rather than mere (in)ability to detect a key violation. Close inspection of stimuli also showed that in two of the melodies used in this experiment, the target tone represented the first occurrence of that specific tone in the melody. If only the preceding context were used to guide the judgement of key violation, the tone could have implied the minor instead of major version of the key, and so may not have been regarded as key-deviant. This may have added to listeners' response uncertainty. In theory, presenting the task stimuli with chord accompaniment may have made more acoustic information available upon which listeners could base their perceptual decisions. However, since the neural activation patterns and associated behavioural outcome resulting from simultaneously presenting multiple tones of different frequencies have, to our knowledge, not yet been determined, chord accompaniment may have introduced an additional variable in the current study design. Despite these methodological concerns, it should be noted, though, that NH listeners still achieved between 85% and 88% correct (for scales/arpeggios and melodies respectively) in key-deviant conditions, whereas CI listeners performed close to chance. This indicates that although both groups' performance was influenced by response uncertainty, NH listeners were affected to a lesser extent than CI listeners.

4.4 GENERAL DISCUSSION

Results show that CI listeners were not able to complete either of the melodic perception tasks presented in this investigation successfully, which suggests that musical context information is not available to CI listeners to the same extent as to NH listeners. The results may be attributed to several causes, notably related to poor pitch resolution afforded by current CI technology. Frequency-to-place mismatch may result in pitch interval being nonlinearly distorted, more limited spectral resolution compared to that available to NH listeners reduces the extent to which different frequencies can be resolved and channel interaction may cause widespread neural activation that would limit extraction of a specific frequency based on the place code for pitch. Together these constraints may present an overwhelming challenge to the auditory processing system with regard to pitch perception of input as required for successful melody perception.

Although the results confirm poor melody perception ability of CI users as evidenced by previous studies, the approach used here may be valuable in disambiguating reasons for such observed behaviour. CI listeners' inability to use the preceding context and syntax markers of an unfolding tone sequence to drive melody perception can possibly be attributed especially to degraded and compromised pitch information being passed on from early feature extraction stages. Since a hierarchical information processing strategy, as has been proposed for music perception (Peretz & Coltheart 2003; Koelsch & Siebel 2005; Brattico et al. 2006), relies on integration of bottom-up information and top-down modulatory influences of especially pitch cues (Stewart et al. 2008; Balaguer-Ballester et al. 2009) In Western tonal music, it is clear that compromised low-level pitch information propagating through the music processing system will reduce the chance of music perception similar to that experienced by NH listeners.

Investigating CI-mediated music perception ability according to an approach that takes the neurocognitive principles of music perception into account provides insights that are not available from studies that consider perception of music-relevant components, such as pitch and rhythm, in isolation. Moreover, the present design goes beyond the conventional investigation of music perception in that it employs unfamiliar melodies to determine whether information that supports generic music listening is conveyed by the implant device.

Results from the studies described in the earlier chapters of this thesis are pertinent to the discussion of the observed outcome. Given that (i) an endogenous mechanism may exist to facilitate better pitch perception than would be expected from the operation of predominantly place pitch, as posited in Chapter 2, and (ii) given that CI users' contextual rhythm perception ability is comparable to that experienced by NH listeners, regardless of additional pitch complexity that may demand extra cognitive resources to be allocated to the stimulus, why can CI users not extract the melodic characteristics of a tone sequence? Or, on the contrary, if tasks that probed perception of melodic characteristics are too difficult for CI users, why is it that some CI listeners do score favourably on closed-set melody recognition tasks? I believe that the answer lies in the use of a closed set. If rhythm cues are the most salient cue reaching the melody processing stage, and it is sufficiently salient to be matched to a memory representation of one of the response options given by the closed, it could well facilitate higher-order cognitive functions to "fill in the gaps", and in a way "reconstruct" the tune sufficiently in the listener's mind to allow a perception of successful recognition. In that sense, music is in the ear of the beholder. However, when the aim is to determine how CI sound processing algorithms can be changed to convey sufficient auditory information to the central auditory processing system so that an unfamiliar melody, which conforms to the general compositional rules of Western tonal music, can be interpreted, the reasons for familiar melody recognition – whether successful or unsuccessful – need to be known. The results of the investigation described in this chapter therefore reflect descriptive music listening ability as a result of using novel melodies in an open-set design, rather than prescriptive music listening as dictated by using a closed-set, familiar melody recognition task.

A measure to evaluate CI listeners' ability to use contextual information during music listening may thus contribute to available test batteries (e.g. Cooper et al. 2008; Kang et al. 2009) by allowing objective and quantitative comparison across different processing strategies, CI user groups and rehabilitation facilities, and tracking of individual user improvement. Given the exploratory nature of the present investigation it should be emphasised that the tasks proposed here are not presented as being an absolute, 'end-of-line' assessment tool, but rather as modules that may, upon further refining and tailoring, contribute to existing test batteries and so provide a more comprehensive assessment of a CI listener's perceptual ability when confronted with music stimuli in a setting akin to everyday listening conditions.

It should also be noted that the task demands may have been too steep for the CI participants and that it may have been difficult for them to have judged, specifically, completion of a melody. However, since the CI participants were all post-lingually deafened listeners, who had all indicated during introductory interviews before commencement of the investigation that they had enjoyed listening to music before having lost their hearing, it was deemed a suitable exploratory approach. The task instruction that a complete melody would likely evoke a sensation of having ended satisfactorily, without the need for additional tones to "let it finish", while an incomplete melody was expected to create a sensation of melodic tension wanting to resolve to a more stable scale tone, was deemed appropriate owing to earlier reports regarding perception of musical expectancy and its violations (also see paragraph 4.2).

Since both tasks used during this investigation assessed listeners' ability to use preceding melodic context during music listening, incorporating a specific one into an existing music perception test battery would depend on the auditory processing level at which perception ability needs to be evaluated. Tasks based on melodic completion may be better suited to gauging listeners' use of global melodic cues, while tests evaluating perception of key violation could provide insight into context-based pitch perception at a local melodic level. It should be noted, however, that the design of the key violation test may have to be improved to accommodate response bias associated with the yes/no design used in this investigation. Furthermore, in future application of the approach explored here, specific attention should be given to (re)designing tasks that are brief enough to be relevant in a clinical setting; in their present form the tasks are too tedious to be incorporated into a music-specific test battery.

4.5 CONCLUSION

This investigation assessed whether CI listeners can use melodic context to aid perception of simple melodies in a setting similar to what would be experienced in real life. Results showed that helpful contextual information is not available to CI users to the same extent as to NH listeners, which points to the pronounced influence of compromised low-level information being propagated through a hierarchical processing system. The approach used during this investigation may contribute to existing measures to assess and improve CI-mediated music perception ability.

Chapter 5

FREQUENCY-DEPENDENT LOUDNESS VARIATION IN SOUND-FIELD LISTENING CONDITIONS

5.1 INTRODUCTION

In electric hearing cochlear conversion of acoustic input to nerve impulses is taken over by direct stimulation of the auditory nerve. The input signal is thus only an approximation of the naturally received stimulus and as such physical stimulus characteristics need to be coded appropriately to achieve approximately natural perceptual experience of the associated physical dimensions. In the case of loudness it is important to apply electrical signals to the electrode array such that the acoustic loudness will be conveyed accurately (Dorman et al. 1993).

The power relationship between loudness and stimulus intensity (Stevens 1955), combined with the absence of the natural compression mechanism of the basilar membrane and refractory control of neurotransmitter release, mean that a small change in applied current can result in a large change in perceived loudness (Moore 2003). The built-in loudness growth function of the speech processor compensates for unnatural loudness growth to some extent, but allows only a much reduced perceptual acoustic dynamic range – between 35 and 45 dB

(McDermott & Sucher 2007; McDermott & Varsavsky 2009) compared to a 60 dB range for speech and music in NH (Zwicker & Fastl 1999).[7]

Stimulus level has been shown to aid cochlear implantees' speech recognition ability (Fu & Shannon 1998; Franck et al. 2002) and although not a critical factor in quiet listening conditions (Fu & Shannon 1998; Fu & Shannon 2000), the benefit of accurately encoded loudness cues may become evident in everyday listening conditions where limited spectral selectivity is available (Shannon et al. 2004, p. 339). Considering further also the perceptual relationship between pitch and loudness (Stevens 1935; Arnoldner et al. 2006), it follows that loudness effects may influence the outcome of psychophysical studies involving frequency-dependent acoustic input such as music.

This context is especially relevant for sound-field studies where a listener's performance in everyday listening conditions, using a clinically assigned processor and settings, is tested. Such testing conditions assume that clinical amplitude mapping (as performed by an audiologist) generates balanced loudness percepts over the frequency range encountered in daily listening. Perceived loudness during CI-mediated hearing is usually controlled by establishing maximum (or comfort (C)) and minimum (or threshold (T)) stimulation levels for each electrode and then applying a single loudness growth function to determine the actual level of stimulation. This ensures that all electrodes are stimulated at a current level that produces an acceptable loudness percept between the T- and C-levels, but does not necessarily mean that each electrode produces an equally loud percept and thus smooth variation of loudness across different electrodes at intermediate stimulation levels may not be guaranteed (Blamey et al. 2000).

Since CI-mediated hearing strives to restore natural sound sensation (Hoth 2007), it is important to quantify the extent to which the perceptual experience of specific stimulus parameters in CI-mediated hearing differs from NH. Furthermore, within the hierarchical framework that underlies music perception and which guided the approach to this study (see

---

[7] These values refer to perceptual acoustic dynamic range and should not be confused with the input acoustic dynamic range, which is set as a processing algorithm parameter.

Figure 1.1, Chapter 1), it is important to be aware of factors that may influence processing at early hierarchical levels and the extent to which their influence propagates to subsequent processing levels. The perceptual association between pitch and loudness and specifically the encoding of loudness cues in electrical hearing, serve as motivation for the work reported on in this chapter. The aim of this investigation was twofold: (i) to investigate loudness estimation behaviour of CI users in sound-field conditions for pure tone signals using their clinically assigned processors and settings and (ii) to compare loudness balancing over a wide frequency range in sound-field conditions with trends observed for NH listeners. The two objectives were addressed in the loudness estimation and loudness balancing tasks respectively.

5.2 METHODS

5.2.1 Subjects

The CI group consisted of eight post-lingually deafened adult users (mean age = 45.6 years) of the Nucleus implant system. All subjects had experienced profound deafness for more than 10 years and had at least two years' experience with their implant device. Seven subjects were implanted with the CI24 array, while one was implanted with the CI22 array. Other than two implantees who used the 3G processor, all subjects used the Freedom processor, fitted with either the ACE or SPEAK speech processing strategy. For three subjects who were bilaterally implanted, only the ear subjectively regarded as the better of the two was used during the experiments. All subjects had participated in earlier research in our laboratory. Other relevant demographic details are provided in Tables A1 and A2 (Appendix A).

Eight age-matched NH listeners formed the control group (mean age = 48.9 years). All subjects achieved audiometric thresholds of 30 dB HL or better at five octave frequencies (250 Hz to 4000 Hz). Four subjects of the control group have also participated in other psychoacoustic research in our laboratory.

All subjects (NH listeners and cochlear implantees) gave written informed consent in line with the requirements of the relevant ethics committee prior to commencement of the investigation. Subjects were compensated for their time at the end of the investigation.

5.2.2 Experimental procedure: Loudness estimation task

All experiments were performed in sound field in a quiet listening environment. CI listeners used their own speech processors and clinically assigned maps as used during everyday listening. Stimuli were presented through a Yamaha MS 101 II loudspeaker placed approximately 1 m in front of the listener on the side of the test ear.

Each listener’s threshold and maximum comfort levels were determined acoustically before commencement of the experiment using a 1 kHz pure tone reference stimulus. The resulting acoustic dynamic range was then divided into 10 equal steps for presentation of a probe stimulus of the same frequency. Twenty probe tokens were presented at each of the ten stimulus levels in random order. Listeners were asked to assign a value between 1 and 100 to describe the loudness of the probe stimulus, with 1 being just noticeable (but certain) and 100 corresponding to the loudest sound that could comfortably be tolerated during the course of the experiment. Responses were submitted via a GUI.

To prevent decision drift over the course of the experiment, listeners were asked to judge the loudness of the probe stimulus relative to that of a reference stimulus. Each presentation round therefore consisted of a reference and a probe stimulus. The initial presentation round presented a 1 kHz reference tone at the acoustical maximum comfort level, which was by default assigned a loudness rating of 100, after which a probe stimulus followed. Listeners used a GUI-based slider bar to assign a rating to the probe stimulus. This probe stimulus, with its associated loudness rating shown on the screen, was then presented at the beginning of the next stimulation round to serve as reference for the next probe stimulus. Two different colours were used to denote presentation of the reference and probe stimuli on the screen.

Reference and probe stimuli were both 1000 ms long and separated by a 1000 ms interstimulus gap. The task was self-paced and a new stimuli pair was not presented before user response had been received. It took approximately 12 minutes to complete the task.

5.2.3 Experimental procedure: Loudness balancing task

The loudness balancing task required listeners to compare the loudness of probe tones at frequencies between 200 and 6100 Hz to that of a 1 kHz reference stimulus. The frequency range for the probe tone was divided into three sets (200–1250 Hz; 1300–2900 Hz; and 3300–6100 Hz), which were presented blockwise. For the three sets, probe tones were selected at 50 Hz, 200 Hz and 400 Hz increments respectively. Listeners had to perform three separate comparison tasks over the course of the experiment, respectively setting the probe equally loud to, just louder and just softer than the reference stimulus. Probe tones were always presented randomly and three repetitions at each frequency had to be completed in each of the comparison tasks. The average comparative loudness at each frequency was subsequently calculated from nine responses (three repetitions x three comparison tasks).

Both tones of a specific reference–probe combination were initially presented at the intensity associated with a listener’s pre-determined comfortable loudness level (70–75% of acoustic dynamic range). The intensity of the probe stimulus had to be adjusted to satisfy the specific task requirement using GUI-based controls. Listeners were allowed to relisten to the stimuli pair after each adjustment and submitted their final judgement only once they were satisfied that the loudness of the probe stimulus compared correctly to that of the reference stimulus. To prevent undue response bias, the probe tone was randomly assigned to either the first or the second presentation slot in a presentation round. This presentation order held for the entire adjustment cycle of the specific reference–probe combination. The reference and probe stimuli were associated with green and red on-screen signals respectively to ensure that listeners correctly aimed their adjustments at the probe stimulus of the pair. Both the reference and the probe stimulus were presented for 350 ms, separated by a 1000 ms interstimulus interval. Listeners completed each block of comparison tasks in approximately 15 minutes.

5.3 RESULTS AND DISCUSSION

5.3.1 Loudness estimation task

Loudness estimation results are shown for CI and NH listeners in Figure 5.1 and Figure 5.2 respectively, with the logarithm of the numerical loudness rating shown on the y-axis and stimulus intensity level shown on the x-axis. NH listeners' behaviour generally followed a linear trend, while CI listeners' loudness growth estimation was generally linear only for part of the dynamic range (except for S4) before levelling off for input levels above 60–70% of the dynamic range.

Linear regression was performed to compare the rate of loudness growth for the two groups. Since CI listeners experience loudness growth only for part of the dynamic range, least squares curve fitting was applied only to the linear section of CI listeners' loudness estimation curves. The transition point where loudness growth entered the plateau phase was defined as the point after which the ratio between at least two consecutive loudness ratings differed by less than 5%. Fitted lines and the associated regression coefficients are shown on each graph. As shown by the fitted lines, CI listeners experienced steeper loudness growth than NH listeners (also see Table 1), which a t-test showed to differ significantly from NH listeners' response curves ($t(14) = 2.71$; $p < 0.02$).

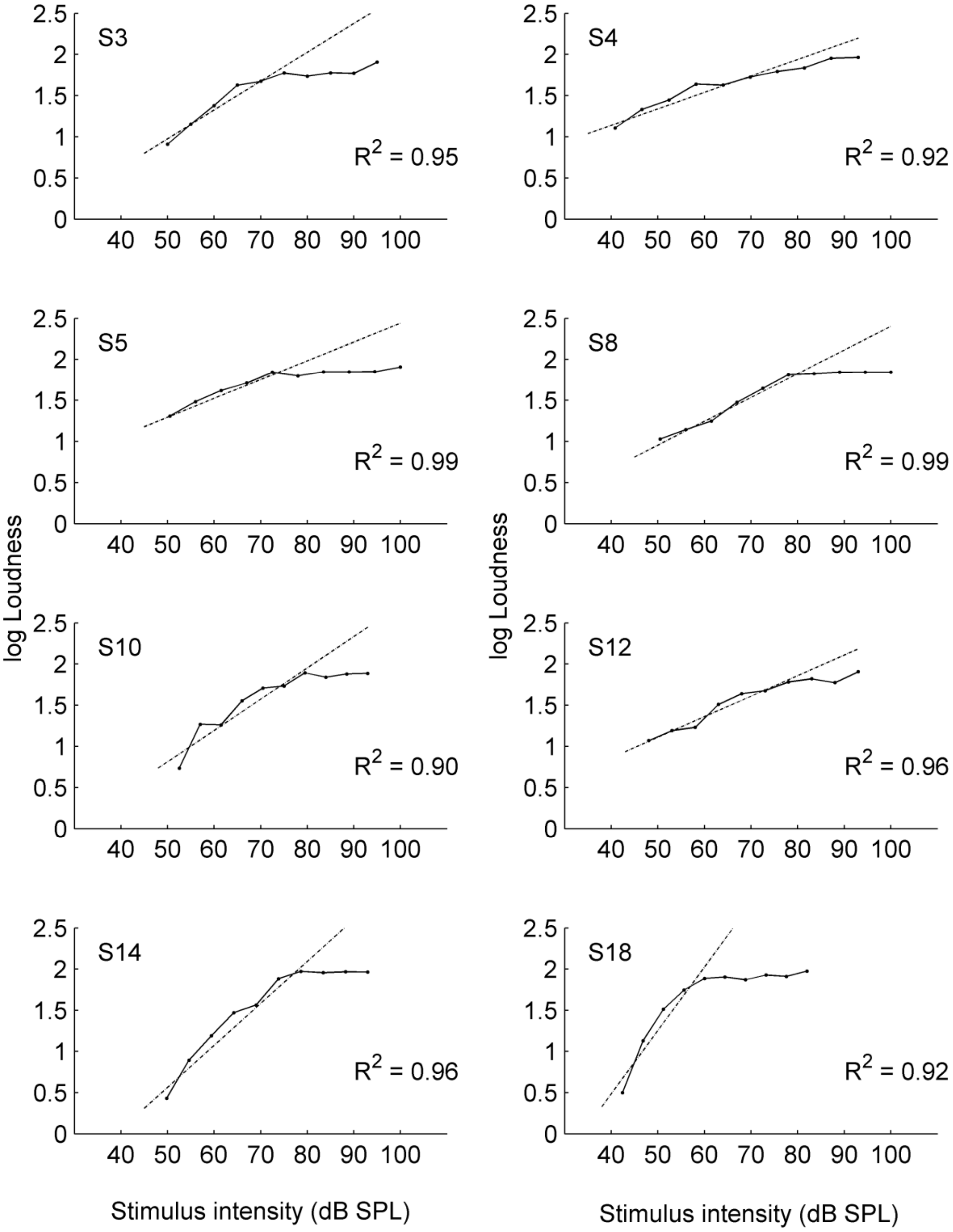

Figure 5.1: Loudness estimation results of individual CI listeners. Regression coefficients for a linear model fit only through the data points corresponding to loudness growth are shown.

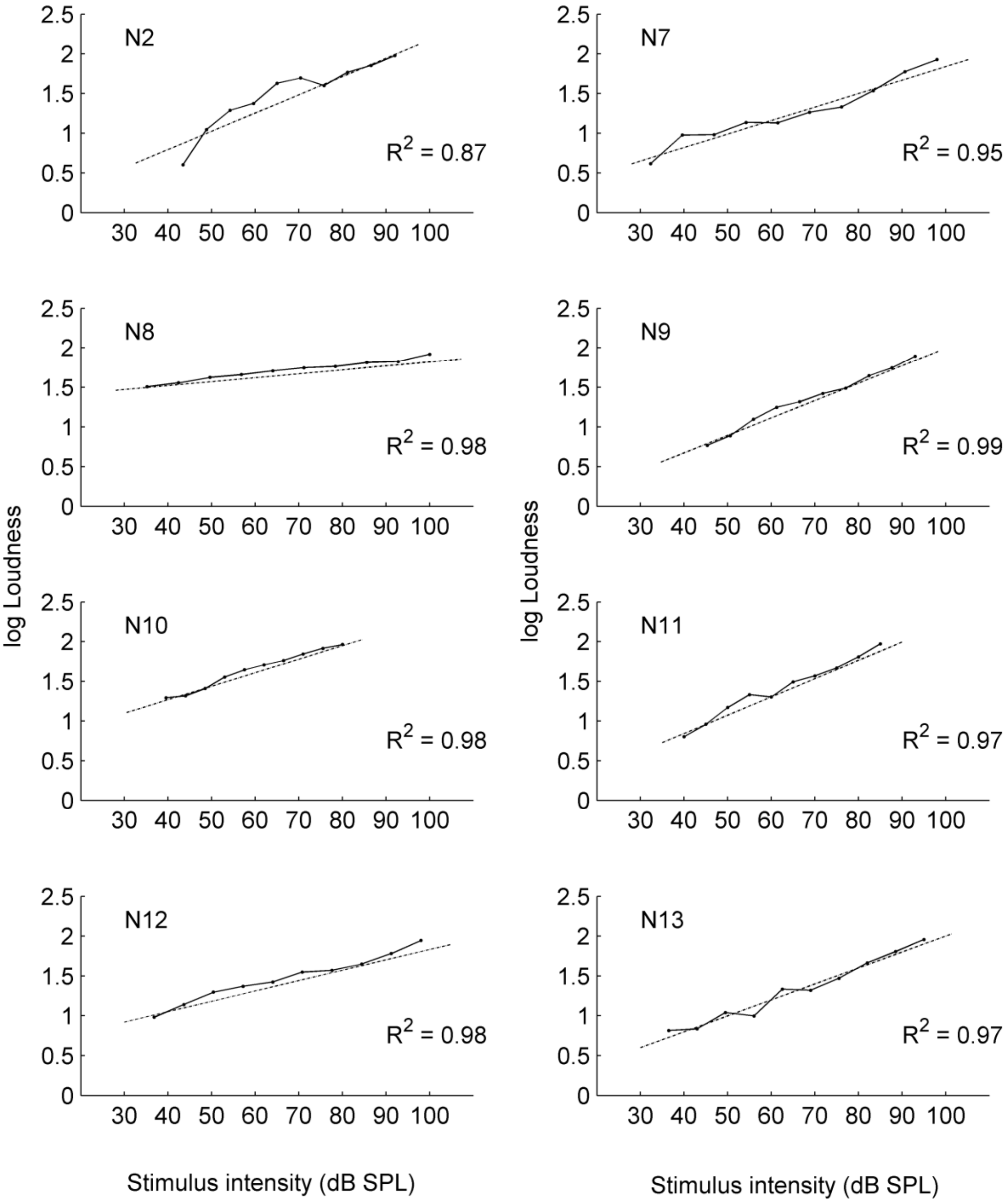

Figure 5.2: Loudness estimation results of individual NH listeners. Regression coefficients for a linear model fit through the data points corresponding to loudness growth are shown.

Table 1: Slopes of linear sections of loudness growth curves of CI and NH listeners and associated regression coefficients. CI users are denoted by S; NH subjects are denoted by N.

Subject	**Slope**	**Regression coefficient**
S3	0.035	0.95
S4	0.019	0.92
S5	0.024	0.99
S8	0.029	0.99
S10	0.038	0.90
S12	0.025	0.96
S14	0.052	0.96
S18	0.077	0.92
N2	0.024	0.99
N7	0.017	0.95
N8	0.006	0.98
N9	0.023	0.87
N10	0.018	0.98
N11	0.024	0.97
N12	0.014	0.98
N13	0.020	0.97

The results are borne out by the average behaviour of the two listener groups, as shown in Figure 5.3.

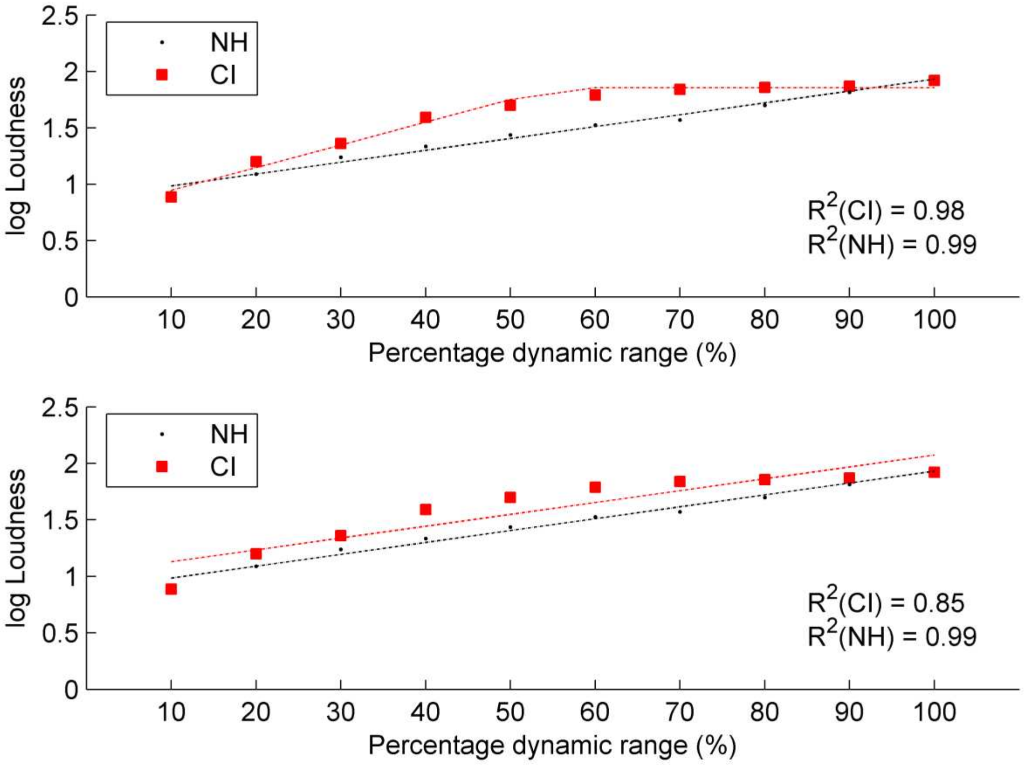

Figure 5.3: Average loudness estimation results for the two listener groups. Since the absolute values of perceptual minima and maxima are not the same for all subjects, data are plotted against percentage dynamic range. The top panel shows CI data fitted according to a two-stage linear model (dashed line), whereas the same data are fitted according to a monotonically increasing linear curve in the bottom panel. For comparison, the NH data set is fitted according to a monotonically increasing linear model in both panels. Regression coefficients for the respective fits are shown in each panel.

Considering the constraints on dynamic range, least squares fitting was performed on average CI data according to

$$f(x) = \begin{cases} ax + c, & x < k \\ c', & x > k \end{cases}$$

with a describing the slope, c the y-intercept of the line with positive slope, k the transition point and c' the y-intercept for the plateau section. Optimised parameter values (as found by using Microsoft Excel Solver function) showed $a = 0.02$ for CI data. Average data from NH listeners were fitted according to $f(x) = ax + c$, with $a = 0.01$ (top panel). The bottom panel shows CI and NH data both fitted to $(x) = ax + c$, with $a = 0.01$. Although the slope of true loudness growth for CI listeners was thus twice that for NH listeners, loudness growth for CI listeners, when considered across the entire available dynamic range, is comparable to that experienced by NH listeners. This may explain the similar loudness balancing behaviour observed for the two groups (see later).

These psychophysical results obtained in sound-field listening conditions are in line with those of McDermott and Sucher (2007), as well as the trends predicted by a computational model of expected acoustic loudness proposed by McDermott and Varsavsky (2009).

5.3.2 Loudness balancing task

The results of the loudness balancing task, averaged over the two groups, are shown in Figure 5.4. Since the absolute reference level was not the same for all subjects (see Methods), results are expressed as the deviation (in dB SPL) from the loudness associated with the reference tone. Relevant musical note names are shown on the secondary x-axis.

The similarity of the loudness balancing behaviour of the two listener groups was evaluated using the Pearson correlation coefficient, similar to the approach used by Throckmorton and Collins (2001). The patterns follow fairly similar trends across the frequency range ($r = 0.65$, $p < 0.0001$), although slightly more variable responses are seen at the higher end of the frequency range. Deviation from the reference level was seldom more than 3 dB SPL, except for the CI listeners at 6100 Hz and NH listeners between 2300 Hz and 3300 Hz. The observation may be attributed to listeners finding the reference balancing method more difficult (Collins & Throckmorton 2000), especially for loudness comparisons between tones at large frequency separations (Lim et al. 1977). The specific pattern of both curves displaying a general peak between 400–500 Hz and a trough between 1100–1200 Hz, suggests an effect of room acoustics.

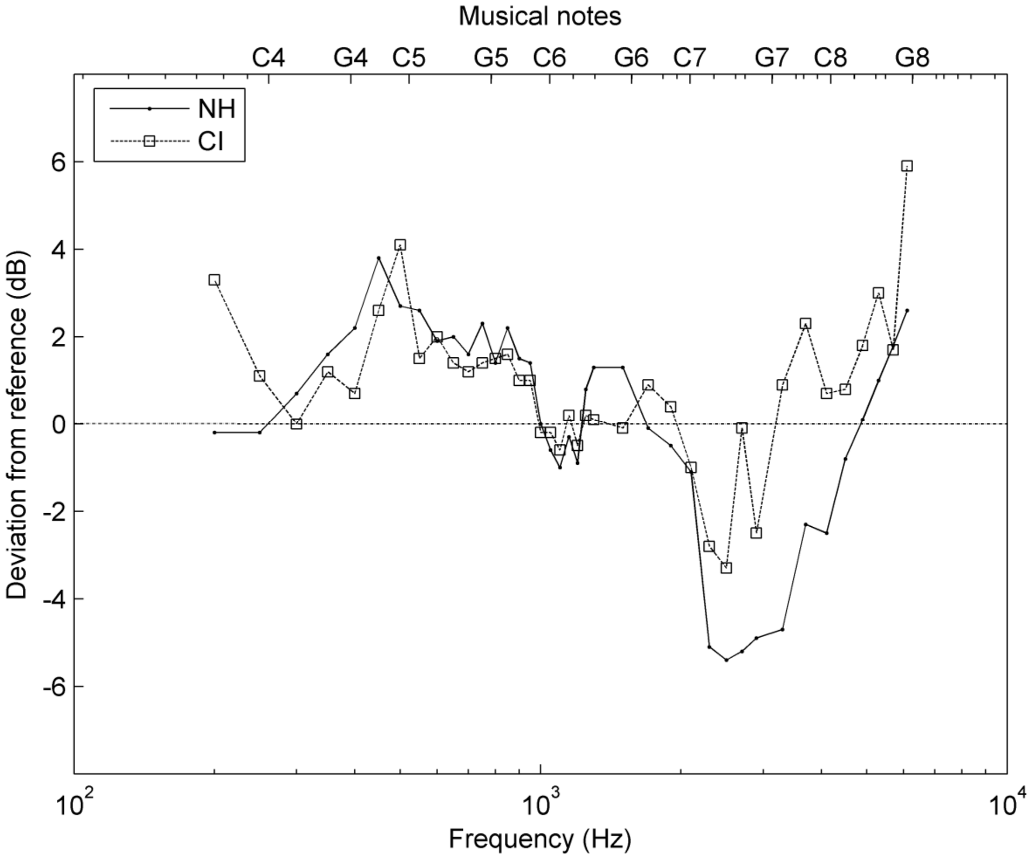

Figure 5.4: Average loudness balancing results for the two listener groups, expressed as deviation from the reference loudness (dashed horizontal line). The frequency range is expressed as musical notes on the secondary x-axis.

The practical relevance of the results, however, needs to be considered in the context of the dynamic range. In line with typical results reported for perceptually relevant listening in everyday conditions (e.g. Fu & Shannon 1998; McDermott & Sucher 2007), NH and CI listener groups in this investigation displayed average acoustic dynamic ranges of 60 dB SPL and 33 dB SPL respectively. (These acoustic dynamic ranges were determined before commencement of the loudness estimation task, see paragraph 5.2.2.) For both listener groups deviations from the reference level thus translate to 10% or less of the acoustic dynamic range, with the exception of the last data point for the CI listeners. Furthermore, for at least 85% of all the successive frequency values used in this study, the deviation differed by less than 2 dB SPL. Vos and Troost (1989) showed that in almost 70% of melodic intervals regularly found in Western tonal music the pitch difference does not exceed two semitones. When the frequency range is thus expressed in musically relevant units, the perceptual effect of loudness deviation is expected to be minimal across a melodic line (consisting of pure tones) presented in sound-field conditions.

5.4 CONCLUSION

The study reported here investigated a group of CI listeners' loudness perception abilities with regard to pure tone signals in sound-field conditions using their clinically assigned processors and compared the results to those of a group of NH listeners. Taken together, the results suggest that CI listeners would experience similar loudness deviation effects as NH listeners, despite reduced acoustic dynamic range and consequently steeper loudness growth curves. It furthermore appears that the loudness deviations experienced by CI users would not adversely affect their listening experience in a frequency-dependent sound-field listening task using pure tone stimuli. The findings do not, however, negate the importance of loudness balancing to rule out loudness as a cue during specific psychophysical tasks (Throckmorton & Collins 2001), but merely suggest that loudness mapping as performed during processor fitting may provide sufficient frequency-dependent loudness balancing for general listening experience (such as music listening) in everyday conditions. It would be interesting to see whether a task using complex harmonic sounds yields similar results.

Chapter 6

GENERAL DISCUSSION AND CONCLUSION

6.1 RESEARCH OVERVIEW

Despite allowing remarkable speech recognition success (Gifford et al. 2008), contemporary multichannel CIs generally afford post-lingually deafened users only limited music perception ability. To date research efforts have investigated CI-mediated music perception ability at two seemingly disconnected perceptual levels, considering perceptual outcomes either after finely controlled feature manipulation at electrode level or after presentation of complete music tokens that include several concurrently varying features. However, a systems approach, which considers electrically mediated peripheral hearing and subsequent central auditory processing together as the link between auditory input and perceptual outcome, may help to shape development of music processing strategies that are applicable to CI listeners' real-world listening conditions.

This study aimed to put forward a systematic approach for investigating cochlear implantees' music perception abilities. By investigating their perception of simple but realistic music-like stimuli in sound-field listening conditions, the study allowed insight into how central processing shapes primary auditory input derived from artificial peripheral stimulation into final behavioural outcome. Although the study still focussed on separate musical features, they were investigated in a musically relevant listening context and interpreted within the framework of the implant's processing capabilities, and so the investigation represents a bridge between understanding the perception of finely controlled elementary musical features and "whole" music listening. Findings from this and similar experimental investigations may contribute to the development of music-specific test batteries that can ultimately help to develop improved music processing strategies.

This study does not cover the entire music listening experience of a CI user, but rather presents an initial exploration into the perceptual outcome brought about by electrical hearing in real-world music listening conditions. Rather than adopting a progressive approach where the findings of a prior analysis serve as starting point for a subsequent investigation, the study focused on four separate aspects that may contribute to understanding CI-mediated music perception. Owing to the importance of successful pitch perception during music listening (Zatorre 2001; Peretz 2002; Peretz & Hyde 2003; Foxton et al. 2004), investigation of CI users' sound-field frequency discrimination abilities (Chapter 2) was a fitting starting point for the study. This was followed by an investigation into implantees' rhythm perception abilities when confronted with short tonal sequences of different rhythmic and pitch complexity (Chapter 3), based on an earlier report (Foxton et al. 2006) about rhythm perception difficulties experienced by tone-deaf listeners. Frequency-related loudness deviation as may be applicable to electrical hearing was also investigated (Chapter 5). Although loudness is not a primary contributor to the musical percept, unbalanced loudness cues associated with different pitch sensations as heard during an unfolding melody may influence a CI user's overall music listening experience.

The work described in Chapters 2, 3 and 5 examined CI users' perception of music-relevant stimuli at a feature analysis level (Griffiths 2003). It was, however, also relevant to explore how, in view of a modular music processing architecture (Peretz & Coltheart 2003), information from early analysis stages affect later perceptual stages. To this end, CI users' ability to infer musical meaning from the preceding context in a novel, single-voice melody with covarying pitch and rhythm cues was investigated (Chapter 4).

6.2 SUMMARY OF RESULTS AND RESEARCH CONTRIBUTIONS

6.2.1 Sound-field frequency discrimination abilities of CI users

In Chapter 2 the frequency discrimination abilities of CI users confronted with pure tone signals in sound-field listening conditions was tested. The investigation aimed, firstly, to put forward typical frequency discrimination threshold values for sound-field listening conditions (since such values had not been available at the time of the study) and secondly, to explain pitch perception outcomes in the context of the filter frequency response curves of the specific speech processing strategy (Research objective 1). Results generally showed that finer frequency resolution than would be expected in sound-field listening conditions may be available to CI listeners; the participants in this investigation were regularly able to discriminate between pure tone frequencies within a single filter band. The frequency discrimination behaviour furthermore appeared to have been influenced by the position of the reference frequency relative to the applicable filter frequency response curve.

The observed frequency discrimination resolution was somewhat unexpected, especially since pitch perception under the specific experimental conditions would predominantly be governed by the place pitch mechanism. Taken together, the findings are thought to point to a possible differential neural activation mechanism, which, based on the current distribution pattern applied to the recruited neural population, may contribute to conveying the frequency information. It was further particularly satisfying to find that such a mechanism may be activated by processor output in sound-field listening conditions without any deliberate attempt to manipulate electrode stimulation, as it suggests that the effect of such a mechanism may translate to perceptual outcome in the context of the complete electrically stimulated auditory system. Such a finding underlines the relevance of a systems approach whereby perceptual abilities of CI users are assessed in listening conditions relevant to their daily auditory environment.

6.2.2 Rhythm perception by cochlear implantees in conditions of varying pitch

Chapter 3 focused on the rhythm perception abilities of CI users when confronted with short tonal sequences of varying rhythmic and pitch pattern complexity (Research objective 2). The investigation was prompted by earlier findings, which showed that listeners who suffer from congenital amusia also experience rhythm perception difficulties (Foxton et al. 2006). These results were surprising in view of earlier observations of selective sparing of rhythm perception abilities when an impairment of the pitch processing stream was experienced (Hyde & Peretz 2004) and independent processing modules being proposed for pitch- and time-based musical relations (Zatorre 2001; Zatorre & Belin 2001; Peretz & Coltheart 2003; Peretz & Zatorre 2005).

Since CI listeners can be regarded as functionally tone-deaf (amusic) owing to limited, and possibly confounded, pitch processing abilities, it was possible that they might experience similar rhythm perception deficits as amusic listeners when confronted with tone sequences of covarying pitch and rhythm complexity (as would be heard during real-world music listening). Some ambiguity regarding cochlear implantees’ rhythm perception abilities does indeed exist (see discussion in Chapter 3) and it was therefore deemed worthwhile, especially within the systems-based approach strived for in this study, to investigate CI-mediated rhythm perception using a similar approach as the one adopted by Foxton et al. (2006).

The results of the investigation showed that CI and NH listeners performed similarly on all the rhythm perception tasks, and specifically that additional pitch complexity did not adversely affect CI listeners’ rhythm perception ability. In addition, CI users exhibited favourable rhythm perception ability regardless of the position of temporal irregularity within anisochronous tone sequences. Taken together, the results suggest that at least for temporal information, input derived from electrical stimulation at the auditory periphery supports adequate central processing to facilitate behavioural outcome comparable to that of NH listeners. Considering the results as the outcome of a hierarchical processing system, and subsequently drawing on findings regarding cortical activation patterns elicited by time-based auditory stimuli during interpretation, contemporary understanding of CI-mediated rhythm perception is extended. Such insights may help to shape efforts to develop music-specific signal processing strategies that are relevant to real-world listening conditions.

6.2.3 Perception of some melodic characteristics by CI users

Listeners’ judgement of syntactic congruency was evaluated in Chapter 4 to investigate whether contextual information yielded by an unfolding melody is available to CI users during music listening. The investigation was developed on the strength of the notion that people have an innate ability to judge melodic endings as complete or incomplete and identify violation of the melodic key, based on the musical context provided by the preceding notes. This approach circumvents the need for closed-set recognition tasks and so frees the perceptual outcome from being influenced by memory representation. Results from such perceptual tasks can thus yield insight into the extent to which the information delivered by electrical stimulation can facilitate general music listening success in real-world listening conditions (Research objective 3).

The investigation comprised two experiments as alternative windows into investigating contextual perception of melodic lines as facilitated by electrically delivered auditory information. Perception of melodic completion is thought to be the outcome of processing effected at an advanced hierarchical level (Brattico et al. 2006) and if CI listeners were able to distinguish between complete and incomplete melodic endings correctly, it could indicate that the higher auditory system is able to assimilate sufficient contextual information for successful music listening, despite unnatural electrical stimulation at the periphery.

Implantees performed close to chance or below for all but the dominant–tonic final progressions, regardless of additional contextual cues such as final note duration. NH listeners achieved between 83% and 96% success on all tasks. The result may point to inadequate pitch analysis experienced by CI listeners at feature extraction level (as described in Chapter 2) creating a substantial perceptual impairment during higher processing stages. However, the finding that the nature of tonic–dominant final progressions was poorly identified while dominant–tonic final progressions appeared to be easier to interpret, indicates that perceptual uncertainty may compound perceptual challenges already presented by insufficient pitch perception ability. The finding that temporal markers provided insufficient contextual support to facilitate perception of melodic completion suggests that within the context of a processing system of which early input is constrained, as is operating in the case of a CI listener, the constraints imposed by insufficient resolution in one dimension overwhelms adequate perception (akin to NH ability) of a complementary stimulus dimension during perception of Western tonal melodies. Understanding CI-mediated hearing as the outcome of an intricate processing system in which several processing modules function in parallel and so is subject to forward feeding as well as modulatory feedback, rather than just as the function of peripheral processing of which the output feeds into individual processing modules, furthers the understanding of the influence early-level processing constraints may impose on eventual perceptual outcome. Although CI users' rhythm perception of tone sequences was similar to that of NH listeners' ability when the task was specifically aimed at perceiving temporal parameters, as shown in Chapter 3, such perception success was not powerful enough to support perception of melodic character of tone sequences where pitch and temporal cues were presented simultaneously as in real-world melodies.

Experiment 2 sought to establish whether CI listeners were able to establish adequate pitch relations between tones of an unfolding tone sequence for supporting identification of key violations. Identification of key violation represents processing at an earlier stage of contextual perception than higher-order syntactic congruency judgement (such as during a melodic completion task) (Koelsch & Siebel 2005). Results showed that sufficient melodic context could not be built up when listening to any of the presented tone sequences, regardless of attention-guiding accent tones. The findings pointed to a sustained effect of inadequate early-level pitch processing during electrical hearing.

Taken together, the results from these melodic listening tasks confirmed that cochlear implantees were not able to use contextual musical information to the same extent as NH listeners. Such inability may stem for compromised pitch perception that originates already at early feature extraction level. When the auditory system is considered as being hierarchically organised, the effect of early processing deficits at integratory processing stages becomes evident. Considering CI-mediated hearing with respect to underlying neurocognitive mechanisms may thus contribute to an improved understanding of the mechanisms affecting CI-mediated music perception.

6.2.4 Frequency-dependent loudness variation in sound field listening conditions

In electrical hearing, stimulus loudness is encoded as the amount of current applied to an electrode (Moore 2003). Owing to the exponential relationship between applied current and perceived loudness (Chatterjee 1999), a slight increase in amplitude may lead to a substantial change in perceived loudness. Further, given the known relationship between pitch and loudness perception (Stevens 1935; Arnoldner et al. 2006) and CI listeners' already limited pitch perception ability, uneven loudness percepts experienced for different frequencies may further worsen their music listening experience (Research objective 4). Loudness in CIs is usually controlled by establishing threshold and comfort levels associated with each electrode and then applying a single loudness growth function to determine the stimulation level on each electrode (Blamey et al. 2000). However, it is possible that controlling loudness in this way may not produce equal loudness across frequencies associated with the different electrodes. In order to put forward a realistic assessment of factors that may contribute to CI listeners' unsatisfactory music listening experience, it is thus pertinent to establish the variation experienced for frequency-dependent loudness perception in sound-field listening conditions.

Results showed that although CI listeners experienced steeper loudness growth than NH listeners, the two groups experienced similar loudness deviation across a frequency range spanning almost five octaves. In the context of the frequency variations in typical Western tonal melodies (for a discussion, see Vos & Troost 1989), such loudness deviation should have a minimal effect during music listening. The findings thus suggest that loudness mapping as performed during processor fitting provides sufficient frequency-dependent loudness balancing for satisfactory general listening.

6.3 GENERAL DISCUSSION

This study put forward a systematic approach to guide investigation into CI-mediated music perception. As outlined in Figure 6.1, the approach builds on the modular information processing scheme underlying the neurocognition of music perception as proposed by Peretz and Coltheart (2003) and findings from various psychoacoustic studies.

Figure 6.1 shows that auditory input is separated along two analysis streams, which deal with pitch and temporal information respectively, already at an early stage of central auditory processing. Within each stream distinct modules for handling separate characteristics of the auditory input exist. However, despite their apparent independence, these modules contribute collectively to processing of pitch and temporal information, and the resulting output combines to drive melody perception at a subsequent processing level. From there output, in turn, feeds into yet higher cognitive association processing modules to eventually create the aesthetically pleasing entity we perceive as music.

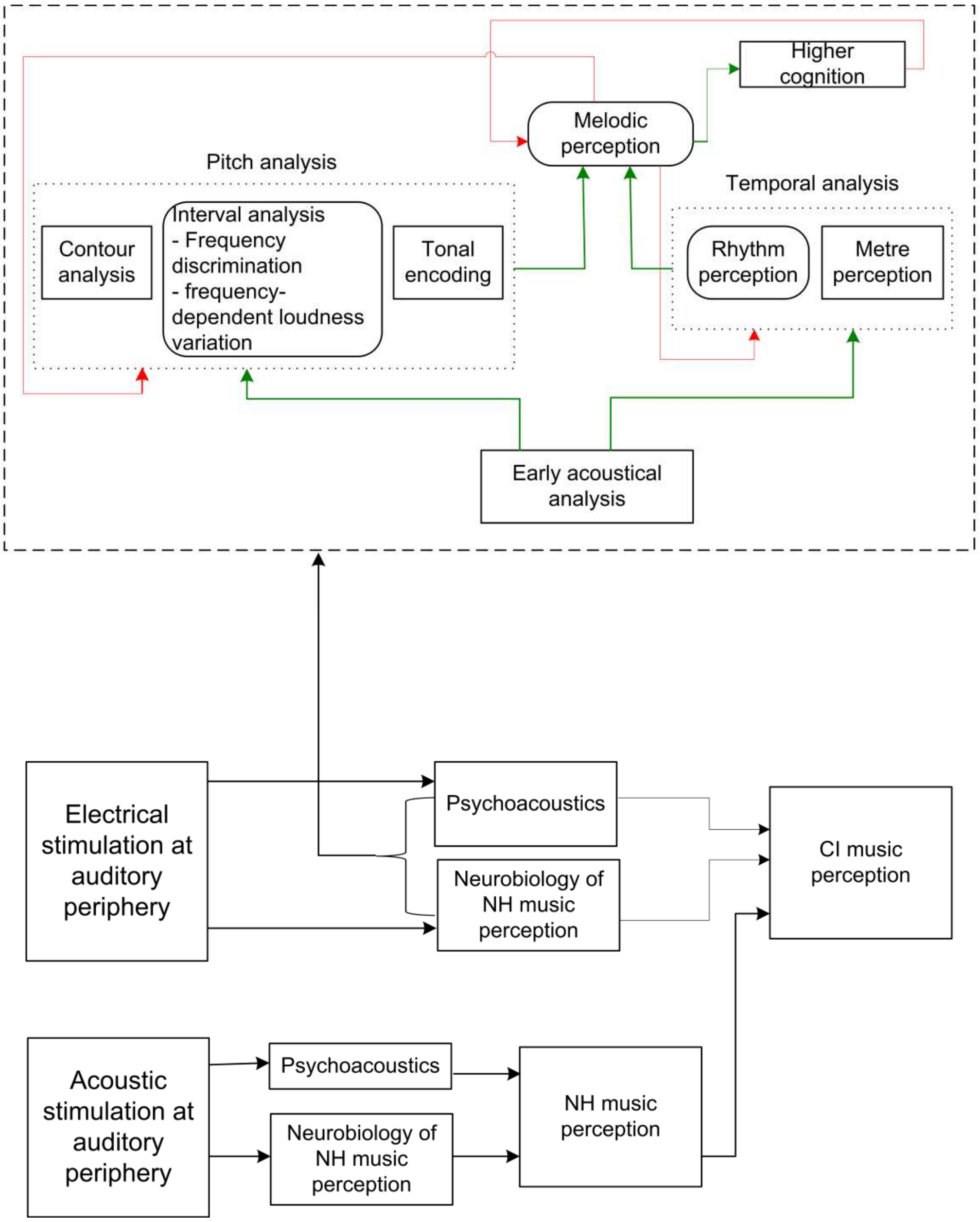

Figure 6.1: Schematic representation of the context of the study (repeat of Figure 1.1). The approach to improve understanding of CI music perception is informed by knowledge regarding psychoacoustics and neurobiology underlying music perception. The neurobiological framework on which psychoacoustic 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.

To date CI-mediated music perception has not been investigated within the framework of a hierarchically organised perceptual system, nor has the outcome of sound-field listening been considered in this scheme. However, using a systematic, hierarchically organised approach to guide psychoacoustical experiments as proposed here may produce valuable insights to understanding CI-mediated music perception by allowing peripheral signal processing to be linked to eventual perceptual outcome.

This study investigated CI users' perception of musically relevant stimuli presented as isolated cues at early analysis levels as well as in combined form at perceptually higher levels in sound-field listening conditions. Although sound-field stimulus presentation is often regarded as allowing the experimenter too little control for interpreting perceptual outcome, implementing sound-field presentation according to an approach informed by the underlying systematics of music perception in NH in this study allowed insights that may not have been found otherwise. For example, investigation into sound-field frequency discrimination (Chapter 2) showed that CI users may achieve finer frequency resolution than a position-governed pitch-coding mechanism would be expected to make available Interpretation of results with consideration to the filter properties associated with the electrodes suggested that stimulation of the auditory nerve as occurs during sound-field listening may activate a differential neural activation mechanism to help signal frequency cues despite pitch-coding constraints imposed by tonotopic mismatch (Fu & Shannon 2002), unfocused neural targeting (Hughes & Abbas 2006) and limited spectral resolution (Shannon 2005).

Similarly, studying CI users' rhythm perception in the context of short tone sequences presented in sound field (Chapter 3) reaches beyond inferring CI users' perception of time-based cues during music listening from their performance on isolated rhythm perception tasks. Results from this investigation showed that CI listeners are able to perceive rhythmic patterns and irregularities equally well as NH listeners, irrespective of tonal information presented simultaneously. The finding suggests that CI-mediated transfer of temporal information perception is adequate to support rhythm perception as necessary during simple music listening. Furthermore, by considering the results against the background of rhythm perception difficulties experienced by amusic listeners whose amusia stems from pitch processing difficulties, it was concluded that the processing stage at which the pitch processing deficit occurs may influence rhythm perception. Within the framework proposed in Figure 6.1 it is thus possible that a central pitch processing impairment may silence the contribution of earlier extracted temporal information during music perception, while CI-mediated temporal information may proceed to later analysis stages unhindered owing to pitch and temporal cues already being separated at the auditory periphery.

Yet, despite finer pitch resolution being possible during electrically mediated hearing and transfer of temporal information allowing satisfactory rhythm perception, the investigation into CI users' ability at a higher perceptual level (where they had to perceive the melodic character of a simple, single-voice melody) shows that current implant technology supports neither satisfactory melody perception nor useful judgement of musical character (Chapter 4). When CI listeners were presented with stimuli that included covarying pitch and rhythm cues in tone sequences that obeyed the theoretical principles of Western tonal music, they were unable to perceive the melodic characteristics musically untrained NH listeners were able to interpret with ease. Findings from neurological studies (Koelsch et al. 2004; Limb et al. 2010), however, show that CI listeners' cortical processing structures are similarly susceptible to processing musical information as their NH counterparts'. When results from the frequency discrimination and rhythmic perception investigations (Chapters 2 and 3 respectively) are thus considered together with findings from the melodic perception study, these point towards inadequate and inaccurate early pitch processing exerting its effect well into higher cortical processing stages. Thus although this investigation confirms that limited pitch processing ability is central to CI users' poor music perception, it presents a context in which a systems-effected perceptual outcome allows signal processing strategies to be adjusted more efficiently. This may allow for either more objective assessment of user performance over time or better comparative studies regarding different signal processing strategies.

In Chapter 5 frequency-dependent loudness variation was investigated. Although loudness is not a primary contributor to music perception, stimulation of cochlear electrodes controlled only by signal processor output in sound-field listening conditions may result in uneven loudness cues being associated across the tonotopically arranged electrodes. The investigation showed that frequency-dependent loudness variation resulting from electrical stimulation is similar to that experienced by NH listeners under the same listening conditions. The finding suggests that loudness settings as set during conventional mapping procedures allow adequately even loudness percepts across a wide frequency range and therefore loudness cues as delivered during sound-field listening are unlikely to significantly confound pitch perception in everyday listening conditions. The outcome may be useful for design of music perception test strategies in that overly rigorous and time-consuming loudness balancing may not be necessary.

6.4 CONCLUSION

6.4.1 Concluding remarks

This study put forward an exploratory investigation into CI-mediated music perception in listening conditions relevant to those an implant user would generally encounter. It drew on knowledge regarding the neurocognition of music perception to guide experimental investigation into how electrical stimulation at the auditory periphery shapes eventual perceptual outcome. Using a systems-based approach, it was possible to explore the extent to which compromised transfer of music-relevant information influences processing at different stages, leading to a complete musical percept. Results showed that low-level pitch processing deficits exert a sustained effect on subsequent higher-order processing and thereby severely complicates successful music perception.

Although the respective investigations described in Chapters 2 to 5 addressed aspects of separate processing modules as would be operating during CI-mediated music perception, they are bound together by their being investigated within a system-relevant context. The overarching rationale behind the approach was that speech processor output at the periphery feeds into the central auditory processing system where it functions within a complete perceptual system where output of several processing modules is integrated to effect a whole music percept. As such CI-mediated music perception was regarded as a function not only of peripheral sound processing, but rather as the *result of speech processor output as handled within an integrated, modular and hierarchically organised central processing system*. It was hence deemed applicable to investigate aspects that may contribute to music perception while considering the central processing mechanism, but also to try to relate them to design parameters as implemented in current sound processing strategies since that is the processing level at which outcome of research regarding CI-mediated music perception can be implemented.

In Chapter 2 frequency discrimination abilities of CI users was investigated in listening conditions that resembled those they are confronted with in daily life. Results were in line with those of previous studies (e.g. Gfeller et al. 2007) in that CI users' frequency discrimination abilities were markedly poorer than had been shown for NH listeners previously. However, the investigation delved deeper than only determining typical frequency discrimination thresholds over a range of frequencies by relating the observed behaviour to a characteristic of the frequency response curves as implemented in the filter bank of the sound processing algorithm. When expressed relative to filter width, a significant number of frequency discrimination thresholds were found to be smaller than one filter width. Considering that each filter spans a range of input frequencies but is associated with one electrode (also see Figure 2.5), discrimination between frequencies falling within the same filter band was unexpected, especially given that the place pitch mechanism was deemed to dominate pitch perception under such conditions. By considering the nature of overlap between filter frequency response curves and how it would influence current delivery to electrodes – and hence neural activation patterns that feed into central processing steps – it was satisfying to find that a mechanism that could effect improved pitch perception in CI-mediated hearing may be available in current-generation sound processing algorithms. Had CI-mediated pitch perception thus been considered only as the outcome of peripheral, device-governed processing, without considering its influence on further processing steps, such hidden pitch perception possibility would not have been found. The findings thus may contribute to future efforts to improve pitch perception abilities of CI users despite the constraints of reduced spectral resolution owing to only a limited number of electrodes being available to handle the entire spectrum of cues delivered to the speech processor in everyday listening conditions.

Rhythm perception ability of CI users was investigated in Chapter 3 with specific interest in determining whether good rhythm perception ability, as has been shown in earlier studies, holds when the task is context based. The investigation was prompted by findings of a study by Foxton et al. (2006), which showed that amusic listeners displayed rhythm perception difficulties which likely stemmed from a pitch processing deficit. This was an unexpected finding owing to the early independence of the pitch and temporal processing streams facilitating music perception, as suggested by Peretz and Coltheart (2003). Seeing that CI users can also be regarded as functionally amusic owing to the limited pitch perception ability afforded by the device, it was deemed suitable to determine whether pitch cues presented together with rhythm cues gave rise to similar rhythm perception difficulties in CI users as amusic listeners. The approach was deemed valuable especially considering that several earlier studies specifically aimed at investigating CI users' rhythm perception abilities were conducted without covarying pitch changes within a tone sequence. Results of the present investigation showed that pitch processing deficits did not influence rhythm perception ability of CI users to the same extent as experienced by amusic listeners in the study by Foxton et al. (2006); rather, rhythm perception behaviour under all test conditions appeared similar to that of NH listeners. Interpreting the results with consideration to the (good) temporal resolution afforded by the implant device suggested that the level at which the pitch processing deficit occurs along the auditory processing pathway may influence the extent to which processing of one dimension influences processing of the other, possibly because the good temporal resolution is able to propagate and sustain its effect throughout subsequent processing steps. Had rhythm perception abilities been considered only as a function of peripheral processor output, without consideration to how it functions within later central auditory processing steps, such understanding of systems-level rhythm perception in CI-mediated hearing may not have precipitated.

Chapter 4 described CI users' abilities to perceive some melodic characteristics of novel, single-voice melodies. The rationale for the investigation was that melody perception, specifically in the Western tonal tradition, requires both pitch and rhythm cues to be perceived successfully to allow perception of a unified musical percept. However, owing to earlier studies of CI users' music perception ability often employing familiar melodies and some favourable outcomes being observed, it was deemed appropriate to disambiguate the reason for such recognition success. The rationale was that having to select a response option from a closed set of familiar melodies may activate higher cognitive processing modules such as memory and/or associative emotional affects to contribute to successful recognition. Although drawing on additional information to help one make a perceptual decision is not undesirable in real-world listening, it does not allow the perceptual outcome to be linked to the nature of information delivered to the central auditory processing system by electrical handling of input at the periphery. It was therefore deemed valuable to determine first whether the information reaching the processing stage at which melody perception is effected is sufficiently salient to signal that a tone sequence is melody-like, for if perception of the melodic character of a tone sequence is not possible, successful melody recognition of familiar melodies as observed in earlier studies is likely to be due to use of other supporting perceptual strategies.

Although the perceptual tasks investigated in this thesis were not meant to follow in a progressive sequence, perception of melodic character was deemed an appropriate perceptual task after investigation of perception of separate elementary stimuli. Furthermore, since sound-field frequency discrimination ability was found to be better than expected (Chapter 2) and rhythm perception behaviour within the context of a tone sequence was similar to that of NH listeners (Chapter 3) it was not an unreasonable question to ask whether the output of those respective processing steps could combine to support perception of the melodic character of a tone sequence. Results showed, however, that despite pitch perception being better than would have been expected when considering the dominant place pitch mechanism at play, constraints such as limited spectral resolution, mismatched frequency-to-place mapping and channel interaction imposed on pitch perception ability may prove too overwhelming during the melody perception stage to allow NH-hearing-like rhythm perception ability to make a substantial contribution to perception at this level. The value of the approach of this investigation lies therein that the ability of CI-mediated auditory input to convey melodic character could be investigated in an objective manner, without being confounded by contribution of additional information derived from higher-order processing modules that form part of the music perception system.

The approach implemented in the study described in Chapter 4 was exploratory and as such the same CI listener group did not participate in each investigation (also see Appendix A and Table A1). Reflection after the investigation had been completed showed that a different study design may have been useful. It is possible that the tasks were too difficult for CI listeners, especially given their pitch perception difficulties and some may argue that such an outcome could have been foreseen. However, since the rationale underlying the entire study was that eventual perceptual outcome is not only the “linear” sum of the outcome of separate processing steps, the investigation was deemed to be justified in the overarching quest of improving current understanding of factors that contribute to (or distract from) successful CI-mediated music perception.

The choice of stimuli may also have contributed to the difficulty of the tasks and it may be argued that using stimuli which included harmonic accompaniment could have provided listeners with more contextual information upon which to base their decisions. However, to deduce processing of peripheral input based only on eventual perceptual outcome means that the researcher sacrifices some experimental control over variables. As applied to CI-mediated perception it implies that auditory input which delivered several frequency components simultaneously may have complicated interpretation of results, since it would not have been known to what extent overlapping neural activation, frequency mismatch or nonlinear frequency distortion may have influenced eventual perceptual outcome. In view of such a design constraint it was thus deemed relevant to attempt the investigation in a rather simplistic way, which, admittedly, may not be realistic for true music listening situations, but which could provide some elementary insight upon which further investigations can build.

In hindsight, designing investigations described in the respective chapters such that the outcomes could have been correlated statistically may have helped to bind the findings together more strongly and so have contributed to delineating the relative contributions of specific processing modules’ outcome in supporting CI-mediated music perception (also see section 6.4.2). However, despite this limitation, the current study design did allow an alternative view into CI-mediated music perception, which may serve as a departure point for future investigations, specifically with the aim of tracking individual listener improvement or functional perceptual benefit afforded by future-generation processing algorithms.

6.4.2 Future directions

As shown through the earlier discussion, this study was not an exhaustive investigation of all the underlying aspects that contribute to CI-mediated music perception, but considered some that may be pertinent to a better understanding of music perception in electric hearing. As such, further investigation into cochlear implantees' perception of musical characteristics such as timbre, metre and melodic contour, performed in listening conditions that approach those an implant user would generally encounter, may further shape understanding of how to translate electric auditory input to useful, behaviourally relevant perceptual output. Especially relevant may be combining investigation of sound-field perceptual ability of CI users with imaging studies in order to form a more comprehensive picture of how the central auditory processing system handles electrically mediated input originating at the peripheral processing stage.

At a more immediate level, further research may involve the following:

- Similar experiments as performed during this study, but with harmonic complexes or true instrument sounds. Results may show how filter output following extraction of harmonic components activates several neural populations, and how such combined neural activity in turn shapes perceptual outcome.

- Delineating the relative contribution of specific music-relevant input cues. Findings may be useful in developing a quantitative, clinically useful measure of general music perception ability of CI users. Such a quantitative measure may help to compare objectively different processing strategies, track changes in user performance over the course of rehabilitation or evaluate the perceptual effect of specific adjustments to processing algorithms or stimulation procedures. Current test batteries for assessing CI users' music perception abilities (e.g. Cooper et al. 2008; Kang et al. 2009) evaluate perception of musical cues in an isolated manner and connecting the outcomes of the different test modules may be useful in establishing a general view of a CI users' music perception ability.