# AUDITORY GROUPING IN COCHLEAR IMPLANT LISTENERS

HUW RHOS COOPER

Doctor of Philosophy

ASTON UNIVERSITY
Psychology, School of Life and Health Sciences
February 2008

# ASTON UNIVERSITY

# AUDITORY GROUPING IN COCHLEAR IMPLANT LISTENERS

**HUW RHOS COOPER**
**Doctor of Philosophy**

**2008**

## Thesis summary

Auditory grouping has been extensively studied in normal-hearing listeners, encompassing the perceptual organisation of both sequences of sounds and mixtures of concurrent sounds. These processes have been studied little in cochlear implant listeners. They experience particular difficulties listening in the presence of competing sounds; poor or absent sound segregation may contribute to this problem. This thesis describes a series of experiments investigating both sequential and concurrent auditory grouping in implant listeners. Some grouping cues used by normal-hearing listeners should also be available to implant listeners, while others (e.g. fundamental frequency) are unlikely to be useful. As poor spectral resolution may also limit implant listeners' performance, the spread of excitation in the cochlea was assessed using Neural Response Telemetry (NRT) and the results were related to those of the perceptual tasks. Experiment 1 evaluated sequential segregation of alternating tone sequences; no effect of rate or evidence of perceptual ambiguity was found, suggesting that automatic stream segregation had not occurred. Experiment 2 was an electrode pitch-ranking task; some relationship was found between pitch-ranking judgements (especially confidence scores) and reported segregation. Experiment 3 used a temporal discrimination task; this also failed to provide evidence of automatic stream segregation, because no interaction was found between the effects of sequence length and electrode separation. Experiment 4 explored schema-based grouping using interleaved melody discrimination; listeners were not able to segregate targets and distractors based on pitch differences, unless accompanied by substantial level differences. Experiment 5 evaluated concurrent segregation in a task requiring the detection of level changes in individual components of a complex tone. Generally, large changes were needed and abrupt changes were no easier to detect than gradual ones. In experiment 6, NRT testing confirmed substantially overlapping stimulation by intracochlear electrodes. Overall, little or no evidence of auditory grouping by implant listeners was found.

KEY WORDS: Cochlear Implant; Auditory streaming; Auditory grouping

# Acknowledgements

With grateful thanks to:

Brian Roberts, for his endless support, patience, wealth of knowledge and many helpful comments on this thesis;

David Landsberger for indispensible help with programming of the 'Impress' software;

Mike Harris and Liz Moores for advice on statistics;

Louise Craddock for her help in collection of ECAP data;

My wife Gwyn, for her patience, support and understanding over six years;

The Hear and Now Trust for financial support.

*The first three years of the research described in this thesis were carried out in the School of Psychology, University of Birmingham. All data were collected during the period 2001-2007.*

*All research was carried out with appropriate ethical approval (Aston University Ethics Committee project number 05/19).*

*Data reported in this thesis were presented at the British Society of Audiology short papers meetings on experimental studies of hearing and deafness in 2003, 2004, and 2007; also at the Conference on Implantable Auditory Prostheses, Asilomar, 2005, the British Academy of Audiology annual conference, 2006, the British Cochlear Implant Group meeting, April 2007 and in Hearing Research, 2007 (see references for full details)*

# Contents

## Chapter 1: Introduction to Auditory Scene Analysis

## Chapter 2: Cochlear implant listening

**Chapter 3: Experiment 1**

**Chapter 4 : Experiment 2**

**Chapter 5: Experiment 3**

**Chapter 6: Experiment 4**

**Chapter 7: Experiment 5**

**Chapter 8: Experiment 6**

**Chapter 9: Discussion**

APPENDIX 1: EXPERIMENT 1: INSTRUCTIONS

APPENDIX 2: EXPERIMENT 2: INSTRUCTIONS

APPENDIX 3: EXPERIMENT 3: INSTRUCTIONS

APPENDIX 4: EXPERIMENT 4: INSTRUCTIONS

APPENDIX 5: EXPERIMENT 5: INSTRUCTIONS

# Tables and figures

## Tables:

## FIGURES:

# Chapter 1: Introduction to Auditory Scene Analysis

## 1.1 Auditory Scene Analysis: History and Background

This thesis reports a series of experiments that investigate auditory grouping in cochlear implant listeners. Before the experiments and their results are described in more detail, this chapter provides an introduction to the history and background of auditory grouping and scene analysis in normal-hearing and hearing-impaired listeners.

The short paper by Miller and Heise (1950) on the perception of sequences of tones ends with the sentence: ‘If one is willing to postpone the question, “Are these patterns melodies?” there is here a large field for worth-while research’. These words have since turned out to be quite prophetic in their anticipation of a fascinating area of research into what Bregman (1990) has termed Auditory Scene Analysis. Miller and Heise discovered, almost by accident, that when the frequency difference between two rapidly alternating tones was sufficiently large, the pattern heard broke up into two melodies instead of a ‘trill’. They coined the term ’Trill threshold’ to describe the frequency separation between the alternating tones that leads to this effect. Their listeners adjusted the frequency of one of the tones while the other was fixed in frequency, and then reported the point at which they heard the change from a ‘trill’ into two distinct and unrelated tone sequences. Miller and Heise also suggested the results they found with auditory stimuli were analogous to what one might expect with corresponding visual figures, and that the principles of Gestalt psychology (see Butler and McManus, 2000, page 18 for a review) are applicable to the auditory situation- a theme that has been returned to on many occasions since by Bregman and others, and which will be discussed in more detail below.

Although research on the topic had been published previously, Auditory Scene Analysis was defined and explained in comprehensive detail by Bregman (1990). He began by elaborating the concept that the purpose of perception and its evolution is to make use of sensory inputs to create a mental representation of the world around us. Auditory Scene Analysis is the process by which the auditory system takes the complex mixture of sounds that is delivered to our ears from the environment around us and sorts it into 'packages' of acoustic evidence, in which each package has arisen from a distinct source of sound. The auditory system must group the acoustic components it receives so as to recover the original individual sources. The complexity of this task is great, as illustrated by the example Bregman (1990, page 5) provided: 'Imagine you are on the edge of a lake and a friend challenges you to a game. The game is this: your friend digs two narrow channels up from the side of the lake. Each is a few feet long and a few inches wide and they are spaced a few feet apart. Halfway up each one, your friend stretches a handkerchief and fastens it to the sides of the channel. As waves reach the side of the lake they travel up the channels and cause the two handkerchiefs to go into motion. You are allowed to look only at the handkerchiefs and from their motions to answer a series of questions: How many boats are there on the lake and where are they? Which is the most powerful one? Which one is closer? Is the wind blowing? Has any large object been dropped suddenly into the lake?'

This example may sound impossibly difficult but in fact illustrates the task constantly performed by our auditory system; although we are generally unaware of the complex processing that underlies our mental image of the acoustic world around us. In everyday situations much of the auditory environment is unconsciously and automatically organised perceptually so that we are able to attend to those sounds that are immediately important. For example, when walking down a busy street our ears are bombarded with a rich mixture of sounds that could include various forms of traffic noise, voices, music, footsteps etc. Our auditory system analyses this complex soundscape and assigns different sounds to different sources, much of which is ignored and merges into the background. The extent to which unattended 'background' sounds are processed has been the topic of much debate and will be discussed in more detail below.

Distinct from vision, which is concerned mainly with the creation of mental images of objects, hearing is concerned with acoustic 'happenings', and the perceptual unit that represents a single happening is referred to as an auditory stream. An acoustic happening may be made up of more than one sound- for example, a series of footsteps, although containing a series of distinct sounds, is perceptually represented as a single auditory stream. Analysis of the acoustic properties of the footsteps i.e., loudness, timbre, reverberation, pitch etc allows us to perceptually allocate them to a single walker. A general principle of auditory scene analysis is that sounds or sequences of sounds that have common characteristics are grouped together perceptually, related to the fact that they are likely to have emanated from a common source. The principles of Gestalt psychology have an application in this regard. For example, the 'Law of similarity' states that we tend to group similar elements together to form an entity; The 'Law of proximity' says that spatial or temporal closeness of elements leads to them being grouped together and perceived as belonging together; and the 'Law of common fate' states that when elements move in the same direction, we tend to perceive them as a unit.

## 1.2 The 'Cocktail party' problem

In one of the most difficult but important auditory situations, that of listening to one speaker's voice in the presence of competing voices, the auditory system is able to perceptually segregate the voice to which we are attending from the interfering speakers' voices by making use of a range of perceptual properties of that voice i.e. timbre, pitch, location etc. This ability to focus one's attention on a single talker among a mixture of conversations and background noises, ignoring other conversations, has been termed 'The cocktail party effect' and was originally described by Cherry (1953). He performed a series of simple experiments in which listeners were presented with a mixture of two spoken messages via headphones, and were required to listen to and repeat one of the messages while ignoring the other. He showed that listeners were able to separate out messages, although with great difficulty, if the same talker was used for both and no spatial cues were available. If the different speech recordings were presented to separate ears (dichotic listening) each could be identified easily and attention switched from one ear to the other at will. He also showed that listeners showed very little awareness of properties of the speech presented to the unattended ear; for example, they were able to report the gender of the speaker but were unaware if a different language (e.g. German instead of English) was presented, or if the speech was played backwards. This implied that very little processing of unattended auditory streams occurs. However, the 'cocktail party effect' in which attention can be momentarily switched to an unattended speaker as a result of hearing a 'primed' word such as the listener's name suggests that at least some processing occurs. Since Cherry's paper, the perceptual basis of this phenomenon has been widely investigated. For example, Brokx and Nooteboom (1982) demonstrated, using resynthesised speech, that constant differences in fundamental frequency (F0) between interfering and target speech messages, presented simultaneously, could be used to improve the intelligibility of the target message. They also showed that differences in intonation (e.g. between a monotonous intonation and a normal intonation pattern) between target and interfering speech provided an additional cue that enhanced the beneficial effects of an intensity difference. This finding was confirmed by Bird and Darwin (1998), who reported two experiments modelled on those of Brokx and Nooteboom (1982) but using stimuli that were entirely voiced and with few stop consonants, with the intention of maximising the effects of F0 differences on intelligibility.

Hawley et al. (2004) summarised four of the cues that contribute to the solution of the cocktail-party problem, namely: i) spatial separation of the target speech from interfering speech (the 'head shadow' effect); ii) differences in the temporal properties of the target speech from the interferers; iii) differences in F0 and iv) differences in linguistic content between target and interfering speech. Each of these cues contributes to varying degrees. For example, the study reported by Kidd et al. (2005) emphasised the role of focused attention along the spatial dimension in multi-talker listening situations; they demonstrated the very large advantage of knowing where to listen when listening to a target speech message in the presence of simultaneous but spatially separated messages.

Although reception of speech is clearly one of the most important situations in which the auditory system has to group auditory stimuli according to their sources, the concept and principles of auditory streaming are also familiar in the field of music. Examples of where a single instrument playing alternating high and low tones gives the effect of two instruments playing are well known. In the Baroque period, when a great deal of music was written for a single instrument but with more than one simultaneous melody, this so-called 'melodic fission' was frequently used by composers. Indeed, the ability to listen to an entire symphony orchestra and follow a melody played on one instrument demonstrates the remarkable ability of the human ear to attend to an auditory stream against a background of complex competing sounds. It is therefore perhaps not surprising that much of the early research into the concepts of auditory scene analysis was based on musical notes (e.g. Miller and Heise, 1950).

## 1.3 Basic concepts of auditory scene analysis

Bregman and Campbell (1971, page 244) described the concept of the 'auditory stream', defined as 'a sequence of auditory events whose elements are related perceptually to one another, the stream being segregated perceptually from other co-occurring auditory events'. They had observed the same phenomenon that Miller and Heise (1950) had previously reported and called the perceptual splitting of alternating high and low tones 'auditory stream segregation'. They went on to demonstrate the effect of this property of auditory perception on listeners' judgements of the order of non-speech sounds. Warren et al. (1969) had shown that listeners were unable to accurately report the order of sequences of four sounds (a hiss, a tone, a buzz and the speech sound / i /) presented in a continuous loop, unless the duration of each sound was lengthened to at least 700 ms. The hypothesis proposed by Bregman and Campbell was that this inability was caused by rapid sequences of differing sounds splitting into separate auditory streams; each of the sounds was perceptually grouped with its own prior and subsequent repetitions, rather than with the other three sounds. Their experiment demonstrated this, in a task using sequences of six pure tones at three high and three low frequencies. Listeners reported the order of the tones, which were rapidly presented in various different orders. The three high frequency tones and three low frequency tones were considered by the authors to be 'within streams'. They found that within-stream judgements were highly accurate, while across-stream judgements were at chance level; listeners were unable to switch their attention from stream to stream fast enough to make accurate order judgements. Based on this, they proposed that the purpose of the organisational process that creates auditory streams is to 'pre organise' sounds and extract higher order perceptual attributes.

The concept that the perceptual properties of sounds are computed *within* and not *across* auditory streams is one that has since been studied and utilised to great effect. This makes sense in terms of our perceptual analysis of the acoustic environment: if sounds within an auditory stream are assumed to emanate from one source, and sounds in a different stream from another, then their individual properties are assumed to arise from accidental co-occurrence. For example, Massaro (1977) showed in a task that required listeners to count sequences of tones that were either at the same frequency or which alternated in frequency, that counting performance was severely disrupted if the frequency separation was over an octave, compared with a two-semitone separation. Recently, Gaudrain et al. (2007) described a task in which listeners had to report the order of vowels in a sequence. They found that spectral 'smearing' of the vowels increased accurate identification of their order, which they hypothesised was related to a reduced tendency to form separate auditory streams. In this case, stream segregation was detrimental to performance- a principle which can be usefully employed in the study of streaming phenomena and which will be returned to later. They also suggested that for hearing-impaired or cochlear implant listeners, reduced spectral resolution may result in a reduced ability to form separate auditory streams, which would increase the difficulties encountered in multi-talker or challenging 'cocktail party' listening situations.

In an example of a task in which stream segregation is *helpful* to performance, Dowling (1973) showed that the successful identification of interleaved pairs of melodies is possible while their pitch ranges do not overlap, but very difficult or impossible otherwise, unless another cue such as a difference in timbre is available. This finding is consistent with the principle that sufficient separation of pitch ranges allowed splitting of the two melodies into two distinct auditory streams; the listeners in that experiment reported that they could attend to either one of the melodies, with the other in the background, but not both simultaneously. The importance of the ability to perceptually segregate overlapping melodies in real-life situations was recently highlighted by Lacherez et al. (2007); they reported that concurrent, overlapping alarm melodies used on medical equipment were very difficult or impossible to discriminate by nurses, a finding that the authors attributed to a failure in auditory stream segregation.

Some of the basic principles of auditory stream segregation were investigated by van Noorden (1975). His study began with the concepts of music and the connections between tones in a sequence forming a melody. He elaborated the concept of *temporal coherence*, in which all the tones in a sequence are perceived as being connected into a 'string'. (For consistency with the rest of this thesis, this term will be replaced with the word 'stream'). In contrast, if a sequence of tones splits perceptually into two sub-sequences, temporal coherence has been lost and *fission* has occurred instead. This fission is the same as the 'melodic fission' mentioned above in the musical context and the 'trill threshold' of Miller and Heise, described above, and corresponds to auditory stream segregation. He employed a tone sequence in which pure tones at two different frequencies, A and B, were put together in an ABA...ABA...ABA... sequence. The advantage of this type of sequence is a salient change in rhythm that occurs when perception ‘flips’ from one to two streams. In this case, temporal coherence is heard as a distinctive ‘galloping’ rhythm which disappears with fission, and so the occurrence of either one or the other perceptual state is highly salient. van Noorden pointed out that fission is always heard with large tone intervals (large frequency differences between tones) and temporal coherence always occurs with small intervals; there is also an intermediate range where the listener in effect has a choice to listen either to the stream ABA...ABA or to either the A..A stream or B..B stream. In practice, if listeners are not instructed to listen for one or the other, their perception is likely to ‘flip’ spontaneously and randomly between one percept and the other, just as it does in vision with the vase/face illusion or with the Necker cube (see Butler and McManus, 2000, page 14 for a review). This is an ambiguous line drawing of a cube first described by Louis Necker in 1832, in which the viewer can see one corner of the cube either as being at the front or the back. The perception ‘flips’ randomly between one and the other, and it is only possible to perceive the cube in one orientation or the other at any one time (i.e., it is not possible to see both at simultaneously). This example of multi-stable perception has been studied extensively in visual perception, particularly to explore the understanding of attentional processes and ‘top-down’ and ‘bottom-up’ processing.

van Noorden described two possible attentional sets that would influence perception in this situation: i) 'selective listening' where the listener tries to hear either the A..A or B..B stream or ii) 'comprehensive listening' where he tries to hear all the tones in the sequence together as ABA..ABA. The former will favour fission, while the latter will favour temporal coherence. This led firstly to his definition of the 'temporal coherence boundary' (TCB), which is the boundary between temporal coherence and fission, *when the listener is trying to hear temporal coherence* (i.e., one stream). This boundary is measured in terms of the frequency difference between tones A and B. Secondly, he defined the 'fission boundary' (FB) as the boundary *when the listener is trying to hear fission* (i.e., two streams). At frequency separations above the TCB, there is inevitable fission- it is impossible to hear the sequence as a single ABA..ABA stream. In contrast, at frequency separations lower than the FB, there is inevitable temporal coherence- i.e. it is impossible to hear the sequence as two distinct A..A and B..B streams.

van Noorden measured the TCB and FB in three listeners while varying both the frequency separation between tones A and B and also the speed of presentation of the tones expressed as Tone Repetition Time or TRT, i.e. the time in ms between the onset of one tone and the onset of the next. The TRT used ranged between 60 and 150 ms. The frequency of tone A was swept above and below the frequency of tone B (held constant at 1000 Hz) and the listeners indicated whether they heard the 'gallop' rhythm or fission. At first they were instructed to hold onto the gallop rhythm as long as they could, i.e. their attention was set on comprehensive listening: this provided a measure of the TCB. Secondly, they were instructed to try to follow the A..A string of tones as long as possible, i.e. their attention was set to selective listening: this gave a measure of the FB. A summary of the results is shown in Figure 1.1, below. The main finding was that the FB is relatively constant and independent of TRT, while the TCB is clearly very dependent on TRT: it increases as TRT is increased, or put another way, as the rate of tone presentation is reduced. As can be seen in the figure, the consequence of this is that at lower TRTs (faster presentation rates) there is a greater tendency to hear two separate streams; the TCB converges towards the FB, and the ambiguous region between them in which either percept may be heard is narrow in terms of frequency separation. In contrast, at higher TRTs (slower presentation rate) a much greater frequency separation is required to produce inevitable separation into two streams and there is a much larger ambiguous region. van Noorden went on to measure the TCB and FB using a method of adjustment, in which the listener was given control over the frequency of tone A; this method provided broadly similar results.

In summary, he showed that a) if a listener is trying to listen to one or other of two distinct streams of tones (selective listening), the frequency separation below which this is impossible depends little on the rate at which the tones are heard; and b) when trying to hear a single ABA..ABA stream, the frequency separation above which this becomes impossible is very dependent on the repetition rate. One can conclude from van Noorden's results that when employing a similar listening task: a) across the range of TRTs that he used, there will be a range of frequency separations in which the perceptual experience is ambiguous and it should spontaneously ‘flip’ between hearing one and two streams; and b) the tendency to hear two streams will be strongly influenced by the tone repetition rate. Although van Noorden demonstrated the effect of the attentional set given to listeners on their perception, these predictions should hold even when listeners are given no instructions either way and simply report their spontaneous experience (e.g., Anstis and Saida, 1985).

**Figure 1.1.** llustration of the influence of TRT (Tone Repetition Time) on a) Temporal coherence boundary (upper line) and b) Fission boundary (lower line). From: Temporal Coherence in the Perception of Tone Sequences, by van Noorden (1975). Adapted with the permission of the author.

### 1.4 Sequential vs. Simultaneous grouping

So far, this discussion has focused on stream segregation in sequences of sounds, i.e. *sequential* grouping; as mentioned above, Bregman and Campbell (1971) described the concept of the ‘auditory stream’, defined as ‘a sequence of auditory events whose elements are related perceptually to one another'. Many of the experimental studies investigating this area of perception have concentrated on this type of listening situation. In real life, many of the mixtures of sounds that the process of auditory scene analysis must deal with include sounds that are temporally overlapping. This requires a process of *simultaneous* grouping, where simultaneously occurring sounds are grouped according to their sources. For example, in the 'cocktail party' situation of listening to concurrent speakers' voices, individual voices overlap in time with each other so the problem to be solved is partly one of simultaneous grouping. As Bregman (1990, page 221) pointed out, 'in ordinary listening we have to group acoustic components both sequentially and simultaneously in order to allocate them appropriately to streams that represent individual sources of sound'. Although much of the further discussion below refers to sequential grouping, some studies are concerned with simultaneous sounds and illustrate features of the underlying processes that are relevant to this thesis.

### 1.5 Auditory streaming is cumulative

Another of the key properties of the process of auditory streaming was described by Bregman (1978): the auditory system appears to begin listening to an input with a bias towards hearing it as a single stream, but it gradually accumulates evidence over a period of seconds that may lead to the input being split into sub streams: i.e. auditory stream segregation is cumulative. He stated that: 'Human perceptual systems seem to be biased toward simple perceptions; therefore, evidence may need to be built up before the auditory system is willing to interpret an input as a product of two sources rather than one' (page 381). In his elegant experiment, listeners adjusted the speed of presentation of sequences of pure tones alternating in pitch (two high frequencies, fixed at 784 and 831 Hz and one low, 330 Hz) until they reached a point at which the sequence appeared to split into two distinct streams; this made use of the fact reported by Van Noorden (1975), mentioned above, that increased rate of presentation in alternating tones leads to greater perceptual 'splitting'. Four conditions were created by varying the number of tones packaged between 4 second silences; there were 4, 8 or 16 tones in each 'package' or no silences at all. Bregman showed that as the number of tones presented in each 'package' increased from 4 to 16, the rate required for segregation to be perceived decreased- i.e. more streaming had occurred. In a second experiment, he showed that the duration of the silent period between sequences of tones was also important; the bias towards hearing two distinct streams fell as the length of the silence was increased up to at least four seconds. Similarly, Beauvois and Meddis (1997) reported the time decay of auditory stream biasing, using a task in which an ABAB test sequence of alternating tones (8 repetitions of each AB cycle) was preceded by a 10-second induction sequence of repeated A tones followed by a silent interval of up to 8 seconds. They showed that the proportion of 'segregated' responses for the test sequence fell exponentially as the length of the silent interval increased; on average most of the decay occurred in the first 4 seconds, in agreement with Bregman's (1978) study. Interestingly, they also found a longer time constant for this decay (i.e. the time taken for 63% of the total drop in the proportion of 'segregated' responses to occur) for musicians

(7.8 sec) compared to non-musicians (1.4 sec), which they suggested might have been due to superior auditory grouping abilities acquired through musical experience.

This decay of the tendency for stream segregation is a property that has been studied extensively since its discovery in order to explore the mechanisms of auditory streaming. This length of time also roughly corresponds to the time taken for streaming to build up to its maximum in Bregman's first experiment (Bregman, 1978). This apparent 'sluggishness' in the auditory system contrasts with many other auditory processes, which have timescales of milliseconds or less. Bregman's explanation was that a relatively slow biasing and unbiasing of the streaming process is valuable, because it provides a 'conservative evidence-accumulating process'. The process of auditory scene analysis is tailored to the properties of real-world listening environments; a constant, rapid allocation and re-allocation of sounds heard to different sources would make a mental representation of the auditory environment confusing and unstable; similarly, it would be inappropriate if auditory streams, or the perceptual allocations of sounds to their sources, ceased to exist if the sounds that they are derived from are momentarily inaudible.

Further evidence of the build-up of auditory stream segregation over time (or the decreasing probability of hearing temporal coherence) was reported by Anstis and Saida (1985). In their experiments, subjects listened to 30 second tone sequences alternating in frequency (between 800 and 1200 Hz) and reported continuously if they perceived coherence (i.e. integration) or segregation. Their first experiment showed that the probability of coherence fell throughout the 30-second sequence, with the rate of decrease dependent on the modulation frequency; more rapid frequency alternation (8 cycles per second, c.f. 4 or 2 cycles per second) led to a more rapid onset of a segregated percept. A further experiment showed that the probability of coherence decreased linearly with increases in log tonal interval (consistent with van Noorden's description of the relationship between the Temporal Coherence Boundary and frequency separation already discussed). Their second experiment also showed that the decay of coherence (i.e. build-up of segregation) continued for up to 60 seconds, which was the longest period over which they tested and much longer than the 4 second period described by Bregman (1978). However, the most rapid fall in the probability of coherence occurred in the first 10 seconds, with a shallower decrease thereafter. Although their results were broadly consistent with Bregman's (1978), Anstis and Saida (1985) proposed a somewhat different interpretation. Whereas Bregman had suggested a cognitive process of accumulation of evidence to account for the build-up of segregation over time, they proposed a more peripheral process of the adaptation of frequency-modulation detectors. However, this account was ruled out by Rogers and Bregman (1993), who showed that an induction sequence containing high tones only, presented immediately before a test sequence of HLH-HLH alternating high (H) and low (L) tones, produced a similar amount of segregation in the test sequence as when it was preceded by a sequence of 'galloping' HLH-HLH tones; this result could not have arisen from adaptation of frequency-modulation detectors.

## 1.6 Automatic versus schema based processes in auditory stream segregation

Bregman (1990) made an important distinction between *primitive* or *automatic* stream segregation and schema-based segregation. The former process, he suggested, is unlearned and innate- evolved to provide the auditory system with important information about the acoustic environment automatically and without any cognitive effort. It is also a 'bottom up', obligatory process driven by acoustic inputs; in automatic stream segregation, if two sequences of sounds are sufficiently different perceptually they will split into distinct auditory streams automatically and cannot be heard as a single stream even if the listener tries. The latter, schema-based analysis is learned and is also subject to the effects of attention, whereas primitive stream segregation is often assumed to operate pre-attentively (although this is the topic of some debate, as discussed in section 1.8). An analogy would be to how knowledge of the words and syntactical structure of a language are used to understand that particular language when heard in adverse listening conditions. This can be thought of as a 'top-down' process that requires effortful listening; it does not replace primitive segregation, and Bregman suggested that the two mechanisms complement each other in the overall process of auditory scene analysis. One example cited as evidence of schema-driven selection by Bregman (1990, page 396) is the study by Scheffers (1982), which reported that listeners were able to separate a pair of synthetic vowels with the same F0 and temporal characteristics when presented simultaneously, presumably by comparison with internal templates or schemas (see also Assmann and Summerfield, 1989).

## 1.7 Theories for the underlying basis of auditory grouping

The understanding of the perceptual mechanisms underlying the process of auditory stream segregation and auditory grouping has evolved in the years since it was first described. Hartmann and Johnson (1991) suggested that peripheral channelling is of paramount importance, i.e. that stream segregation is mediated largely by a peripheral mechanism, in which auditory stimuli are processed through separate peripheral channels; this could be based on either frequency (i.e., auditory filters) or on ear of presentation. They investigated listeners' abilities to identify pairs of interleaved melodies (taken at random from a closed set of 12), while varying a number of characteristics between the two melodies including: pitch, ear of presentation, timbre, level, envelope, duration, interaural time difference, added noise, altered rhythm and pure tones vs. 'rough' tones (the latter created by multiplying the pure tones in the melody with a 12.5 Hz tone, resulting in two components separated by 25 Hz and with a 'roughness' of tone). In each case, two melodies A and B were interleaved; melody A was presented on the odd-numbered tones, and melody B on the even-numbered tones. In a 'null' condition, melody A and melody B tones were identical; in each of the others, one of the differences listed above was introduced. The intention was to explore which characteristics of the interleaved melodies could be utilised as cues for stream segregation and so provide better performance than in their 'null' condition in which the tones in each melody had identical characteristics. The key question which they posed was whether stream segregation is possible in the absence of peripheral channelling. They found that those conditions that led to the excitation of different peripheral channels, e.g.

frequency (an octave shift between interleaved melodies) or ear of presentation, led to enhanced performance (i.e. greater stream segregation), compared with those that did not involve differences in peripheral excitation, e.g. temporal envelope (slow vs. fast tone rise time) or 'roughness'. They concluded that 'peripheral channelling is almost an adequate explanation of the stream segregation effect', but also admitted that their data did show evidence of some enhancement in segregation when there was a difference in tone duration between the two melodies, which does not fit with the peripheral-channelling theory. They also demonstrated that separation of sounds into distinct peripheral channels does not *guarantee* stream segregation, but suggested that it is of paramount importance.

Some support for the peripheral channelling hypothesis was supplied by Beauvois and Meddis (1996), who described their computer simulation of the stream segregation of alternating pure-tone sequences. This model was based on simple signal processing analogous to the physiological systems present in the peripheral auditory system. This included subjecting the acoustic signal to a peripheral frequency analysis and splitting it into channels, each with a different band-pass frequency response. The signal was then subject to a series of processes similar to those known to occur in the auditory system (see e.g. Moore, 1995), culminating in a segregation / coherence decision-making stage. They evaluated the performance of the computer model compared with reports on similar 'real life' listening tasks, as discussed above. For example, the interaction between TRT (Tone Repetition Time), frequency separation and the probability of 'coherent' (integrated) outputs was compared with van Noorden's (1975) data. The resulting pattern was consistent with his results, showing the same relationship between the temporal coherence boundary and TRT, and the stability of the fission boundary across a range of TRT's. The model also reproduced the build-up of segregation over time, as described by Bregman (1978) and Anstis and Saida (1985); the auditory system begins with an assumption of coherence, and two sequences of tones are judged to be segregated after about 4 or 5 seconds (a similar timescale to that described by Bregman, 1978). In summary, a model based entirely on peripheral channelling was able to reproduce many of the properties of stream segregation that had been previously described.

However, since Hartmann and Johnson's study and the Beavouis and Meddis (1996) model, there has been a steady accumulation of evidence that stream segregation can occur without differences in peripheral excitation patterns. For example, Singh and Bregman (1997) investigated the contribution of both temporal properties (i.e. envelope: rise/fall times of either 5/95 ms or 95/5 ms) and number of harmonics (2 or 4 harmonics) to stream segregation of complex tones. In their task, listeners reported the point at which an ABA..ABA sequence of alternating complex tones perceptually segregated into two separate auditory streams as the F0 difference between A and B increased in quarter-tone steps. This 'crossover point' provided a measure of the strength of the tendency towards stream segregation; differences between tone A and tone B that enhanced segregation led to the crossover point being reported more quickly, as a smaller F0 difference between A and B was required for two distinct streams to be perceived. They showed that the crossover point for the condition in which A and B had the same number of harmonics but *differed in temporal envelope* was significantly lower than when A and B were identical in terms of number of harmonics and temporal envelope; this demonstrated that envelope differences alone can contribute to stream segregation. Thus, they showed that non-peripheral channelling cues can *enhance* stream segregation, but did not show that they can cause segregation on their own.

Vliegen and Oxenham (1999) reported two experiments in which listeners were able to a) perceive stream segregation in ABA..ABA tone sequences and b) recognise melodies interleaved with distractors, using complex tones, on the basis of differences in fundamental frequency. They showed that in both cases stream segregation was possible with complex tones consisting only of high, unresolved harmonics (harmonic numbers of 10 or higher) with a common passband. Any change in F0 in their experiment would not produce peripheral-channelling cues. They therefore showed that non-peripheral channelling cues can be used to select a sub-stream (voluntary fission), but did not show that they can force *automatic* stream segregation.

Cusack and Roberts (2000) showed that differences in timbre alone could enable selection of a melody from distractors, again with spectral differences being kept to a minimum. Their experiment employed three conditions. In the first condition, pure tones were used for both the target melody and the distractors, and both were in the same frequency range (no difference). In their second condition, the distractors were narrow band noises in the same frequency range as the melody tones (timbral difference). The third condition used distractors that were pure tones in a frequency range one octave higher than the melody tones (frequency difference). They found that performance was best (lowest melodic-alteration threshold) in condition 3 (frequency difference between melody and distractors); this is consistent with the findings of Dowling (1973) already described. Performance was worst in condition 1 (no frequency or timbral difference between melody and distractor). However, the timbral difference alone used in condition 2 gave rise to significantly better performance than the 'no difference' condition. Cusack and Roberts asserted that peripheral channelling by frequency could not account for this finding.

Vliegen et al. (1999) employed a temporal discrimination task to investigate whether segregation based on periodicity information alone occurs when performance on the task is *improved* by integration. This provides an indirect measure of primitive stream segregation, because streaming will lead to impaired performance despite the efforts of the listener. Their task required the detection of a temporal shift (delay) in the onset of tone B in an alternating tone sequence ABA..ABA..ABA.. with frequency intervals between A and B of 1, 4, 8, 13 or 18 semitones. Each sequence consisted of four ABA triplets, flanked by two paired tones A..A. The initial A..A tones were included to encourage the perception of segregation; the overall length of the sequence (2580 ms) was long enough to allow segregation to build up. Deterioration in performance on such a task (higher shift thresholds) is expected with greater frequency differences between stimuli A and B, which should lead to more segregation and so make timing judgements more difficult. Three different types of stimuli were used: a) pure tones, b) complex tones containing high, unresolved harmonics with a varying F0 and fixed spectral shape, and c) complex tones with high, unresolved harmonics with a fixed F0 and varying spectral passband. They found that, as expected, performance worsened with increasing differences between A and B in terms of a) frequency, b) F0 or c) centre frequency of the spectral passband in all three stimulus conditions-including both spectral *and* periodicity–based differences. For example, for pure tones, the shift threshold was around 15 ms for a frequency separation of 18 semitones compared with around 7 ms for a 1-semitone separation. However, the effect found for the varying F0 condition was more modest than for the other two conditions. They concluded that if the decline in performance they recorded reflects stream segregation, then stream segregation based on periodicity alone is at least partly involuntary – i.e. it occurs even if listeners try to hear the tone sequence as a coherent whole.

Roberts et al (2002) pointed out that the results of the studies reported by Vliegen and Oxenham (1999) and Cusack and Roberts (2000), although showing that segregation was possible in the absence of peripheral-channelling cues, were explicable in terms of schema-based selection aswell as primitive stream segregation. In order to investigate whether *primitive*, automatic segregation really can occur in the absence of peripheral-channel cues, they used a temporal discrimination task in which good performance depends on hearing two sequences of sounds as a single stream (integration), as in the task used by Vliegen et al. (1999). They used complex tones containing unresolved harmonics of constant F0; pitch was changed in value (by one octave) and strength by changes in phase, which also affected timbre. This provided cues for stream segregation without any change in spectral content. Listeners had to detect anisochrony heard as an irregularity of rhythm, in sequences of alternating tones, which provided a measure of temporal discrimination. In order to encourage build-up of segregation, the anisochrony was introduced after the first 6 AB cycles of alternating tones, which were entirely regular in rhythm. Differences in phase (providing differences in perceived pitch) gave rise to poorer performance in temporal discrimination thresholds, without any differences in F0 or passband. This provided indirect evidence of primitive stream segregation in the absence of spectral cues. They went on to demonstrate that phase differences could also be used as a cue for subjectively reported stream segregation in a task similar to that used by Anstis and Saida (1985), in which listeners continuously reported over 30 s whether they could hear coherence or segregation. Thus, primitive stream segregation without peripheral-channelling cues was demonstrated using two complementary approaches. However, it should be noted that some of the effects of differences in phase might have reflected loudness differences produced by differences in peripheral compression. Nonetheless, Stainsby et al. (2004) showed that, although such compression effects may have influenced the results, the main findings of Roberts et al (2002) were upheld.

The debate concerning the role of peripheral channelling in auditory stream segregation is ongoing. Much of the evidence so far has suggested that peripheral channelling frequently has a dominant role in mediating streaming effects; however, there is also emerging evidence that it is not a *requirement* for stream segregation, and stream segregation has been shown to occur when successive sounds presented to the same ear have essentially the same excitation patterns (Moore and Gockel, 2002). As discussed, cues to stream segregation that have been identified include:

- Differences in amplitude envelope
- Differences in F0
- Differences in timbre
- Differences in phase (providing differences in perceived pitch and/or timbre for unresolved harmonics)

Moore and Gockel (2002) proposed a hypothesis as follows: the extent to which sequential stream segregation occurs is directly related to the degree of perceptual difference between successive sounds. They suggested that *any* sufficiently salient perceptual difference may lead to stream segregation, regardless of whether or not it involves peripheral channelling.

### 1.8 The role of attention in auditory stream segregation

The role of attention in auditory perceptual organisation has recently been a topic of research and some debate, and is relevant to considerations of the relative roles of automatic and schema-driven processes. A previous assumption was that many of the processes underlying auditory scene analysis were automatic and pre-attentive- i.e. largely uninfluenced by attention. Some evidence that attention might affect the build-up of stream segregation was provided by Carlyon et al. (2001). They used 20-second sequences of repeating ABA..ABA.. tones (as used previously by van

Noorden, 1975, and others) presented to one ear of the listeners, accompanied by distracting noise bursts presented to the contralateral ear. If attention was directed towards the tone sequences throughout, reported stream segregation built up over the first 10 seconds as expected (Bregman, 1978). When the listeners' attention was focused on the noises presented to the opposite ear for the first 10 seconds and only then switched to the tone sequence, the reported segregation was reduced, and similar to that which would have occurred as if the sequence had just started. The authors suggested that this lack of build-up of segregation during the unattended period was evidence that attention is a requirement for it to occur. This result was replicated and extended by the experiments reported by Cusack et al. (2004). They showed firstly that the same effect of switching attention away from a sequence of alternating tones in reducing reported segregation occurred if the distractor noise bursts were presented to the same ear as the tone sequences. Secondly, they showed that the build-up of segregation was reduced even if the distracting noise bursts were from a frequency region that overlapped with that of the alternating tones sequences, instead of from a separate frequency range as in the previous experiments. Thirdly, they reported that a brief (5 second) switch in listeners' attention away from the tone sequences to the distracting noises in the opposite ear had a very similar effect in 'resetting' the reported stream segregation to that produced by a 5 second gap in the sequences, even when the tone sequences were continuous throughout. Their explanation of all these results was that the build-up of segregation of the tone sequences was reset when the noise bursts became the focus of attention. This differs slightly from the account of

Carlyon et al (2001), who proposed that attention was a *requirement* for build-up of stream segregation, rather than a switch in attention as a cause of resetting.

The question of whether diversion of attention to either a visual task, or to a cognitive task requiring no sensory input, would have the same effect as another auditory task on reducing the build-up of auditory stream segregation, was investigated by Carlyon et al. (2003). They reported an experiment in which listeners were presented with 13-second ABA..ABA..tone sequences (tone A fixed at 400 Hz and tone B either four or six semitones higher). During the first 10 seconds of the sequence, listeners were instructed to either a) pay attention to the tones, b) count visual targets presented on a screen in front of them or c) count backwards in threes from a randomly generated number given to them at the start. After 10 seconds they were then instructed to judge whether they could hear one or two streams for the remaining 3 seconds of the ABA..ABA sequence. They found that the number of segregated (two stream) judgements reported was significantly lower in the visual and counting conditions than in the auditory condition; they concluded from this that the build-up of auditory stream segregation can be inhibited by diverting the listener's attention to a competing task in the visual domain, or to an unrelated cognitive task, as well as to a competing auditory task. However, the reduction in segregation produced was smaller than that seen in the previous studies using noise bursts (Carlyon et al. 2001). Also, it should be noted that attentional switching alone could have accounted for these findings; build-up in segregation might have occurred without attention being directed towards the tones, but then have been reset by the switch in attention after the first 10 seconds of the sequence.

As noted by Roberts, et al (2007), the effects of switching attention in resetting the build-up of stream segregation are comparable to the effects of abrupt changes in stimuli. For example, the experiment described by Rogers and Bregman (1998) used high (H) and low (L) pure tones in HLH-HLH sequences (frequency separations between H and L of between 5 and 14 semitones). They measured the frequency separation at which the 'galloping' rhythm disappeared (i.e. the Temporal Coherence Boundary or TCB) in 1.2-s test sequences preceded by 4.8-s induction sequences. They showed that a 'no change' condition, where the induction sequence was identical to the test sequence, produced the lowest TCB (i.e. the greatest build-up of segregation). A control condition where the induction sequence consisted of a burst of white noise produced the least segregation. Changes between the induction and test sequences, i.e. sudden or gradual changes in lateralisation from right to left, also gave rise to reduced segregation in the test sequence; sudden changes produced a stronger effect than gradual ones. They also showed that a sudden *increase* in loudness between the induction and test sequences gave rise to a resetting of the build-up of segregation, while a sudden *decrease* in loudness, or gradual loudness changes, did not.

In seeking to understand and explain the findings of Carlyon et al. (2003), Cusack et al. (2004) and Cusack and Carlyon (2004) point out that selective attention and perceptual grouping differ in character. As they state, selective attention is a top-down process, primarily under conscious control of the listener, requiring knowledge of the characteristics of the target of interest. They proposed a 'hierarchical decomposition model' to account for the interactions between stream segregation and attention (see figure 1.2).

**Figure 1.2** Hierarchical decomposition model. From Cusack et al. (2004).

In this model, unattended groups of auditory inputs are not elaborated. For example, when attention is directed to speech in a background of music and traffic noise, the music is not perceptually fragmented into its component elements; this has the advantage of concentrating processing capacity on the target of interest. This is also consistent with the original finding by Cherry (1953) already mentioned, i.e. in a dichotic listening task in which the listener's attention is focused on speech material in one ear, the listener has very little awareness of the characteristics of speech in the other ear beyond knowing that it is speech and the gender of the speaker. The hierarchical decomposition model is clearly a useful step forward in understanding auditory scene analysis and the inter-relationship between peripheral and central mechanisms.

## 1.9 Auditory stream segregation in hearing-impaired listeners

One of the most commonly reported difficulties for listeners with sensorineural (cochlear) hearing loss is the problem of discriminating speech in noisy situations or in the presence of competing speech. Hearing aids improve audibility by amplifying sub-threshold sounds to within the comfortable listening range, and current-day digital aids employing Wide Dynamic Range Compression achieve this more successfully than in the past (see e.g. Chung et al, 2007). Considerable effort has gone into the design of digital signal processing in currently available aids, with the intention of reducing the noise content of the processed signal and enhancing speech sounds. Features such as adaptive directionality in multiple-microphone aids can help to improve signal-to-noise ratios in difficult listening situations (Luts et al, 2004). However, although performance in noise with the best modern hearing aids is reportedly better than before, the problems experienced in noise by hearing-impaired listeners are still substantial, and greater than for normal-hearing listeners. One explanation for this is the reduced frequency resolution known to result from cochlear hearing loss; widening of auditory filters caused by cochlear damage leads to increased masking of signal by noise. Also, if the account of Beauvois and Meddis (1996) is correct (see above), the reduced frequency selectivity associated with cochlear hearing loss should lead to dramatic changes in the perception of sequences of sounds (see Gaudrain et al, 2007). For example, in a sequence of tones of alternating frequency, greater frequency separation between the tones should be required for the perception of fission as opposed to fusion.

In order to test this prediction, Rose and Moore (1997) investigated sequential auditory stream segregation in bilaterally hearing impaired (HI), unilaterally HI and normally hearing (NH) listeners. The task used ABA...ABA... sequences of alternating tones, in which tone A was at one of four frequencies (250 Hz, 500 Hz, 1 kHz or 2 kHz), and tone B began at a higher frequency than A, and was then swept downwards. Listeners were instructed to hear fission (i.e. a segregated percept) for as long as possible and report when their percept changed to a 'gallop', which was used as a measure of the fission boundary (FB). Rose and Moore showed that for NH listeners, the FB was roughly independent of the frequency of tone A, if expressed in terms of $\Delta$ E; this is the E value for tone B minus the E value for tone A, where E is the number of ERBs or Equivalent Rectangular Bandwidths (see Glasberg and Moore, 1990). This value is about 0.4 $\mathrm{ERB}_N$. They asserted that this finding is consistent with the model of stream segregation proposed by Beauvois and Meddis (1996), if it is assumed that the FB corresponds to the point where peaks in the excitation patterns evoked by successive tones are separated by a constant distance along the basilar membrane. They also found that half of the bilaterally HI listeners had larger $\Delta$E's at the fission boundary relative to the NH listeners (i.e. a larger frequency difference between tone A and tone B). This implied that a larger difference in frequency was required for some HI listeners to form two auditory streams than for NH listeners. However, they also found that for the four unilaterally HI listeners tested, there was no consistent difference in FB between their normal and impaired ears. This is *not* consistent with the theory of Beauvois and Meddis (1996), which would always predict a larger frequency difference between A and B at the FB for the impaired ears than for the normal ears, due to the reduced 'sharpness' of auditory filters and reduced frequency selectivity in ears with a sensorineural hearing loss. Their results therefore suggested that the FB is not *solely* determined by peripheral frequency selectivity, as predicted by the Beauvois and Meddis model.

More recently, Grimault et al. (2001) predicted that reduced frequency resolution would lead to a reduction in the sequential stream segregation of complex tones in hearing-impaired listeners, and that this would help explain the problems of this group when listening to speech in noise. To test this, they investigated performance on a stream segregation task using repeating ABA-ABA sequences, where A and B were complex tones with a common passband (1375-1875 Hz); the fundamental frequency for A was set at either 88 or 250 Hz, while the F0 for B was varied between 88 and 352 Hz. Three groups of listeners participated: young normally hearing (YNH), young hearing impaired (YHI), and elderly hearing impaired (EHI). Listeners reported at the end of a 4-second sequence whether they had heard one or two streams. When the F0 of A was 88 Hz, there was no difference in reported stream segregation (as the F0 of B was varied) between the YNH, YHI or EHI listeners. The explanation given by the authors was that this was because the harmonics of the complex tones were not resolvable by any of the listeners and had a common passband; age or hearing loss had no effect on reports of streaming. Significant degrees of stream segregation occurred with a difference in F0 between A and B of one octave, so that presumably harmonics in the B complex tone were unresolved. With an F0 of 250 Hz, differences between the groups of listeners emerged, and streaming scores were systematically greater for the YNH group than for the EHI group. The authors suggested that this was because the harmonics were resolved in the NH listeners but not in the hearing-impaired group, and that this difference in resolvability led to the difference in streaming scores. This supported the idea that frequency resolution is a key factor in differences in stream segregation for these listeners.

Rose and Moore (2005) further investigated the relationship between frequency difference limens and fission boundaries in normally hearing and hearing-impaired listeners. They first measured the FB using an identical method to that used by Rose and Moore (1997). They found that the FB for normally hearing listeners was generally constant across a range of frequencies from 250 to 8000 Hz, when expressed in ΔE units. They then measured the frequency difference limens (FDLs) in the same listeners at the same frequencies as used for tone A in the streaming task. FDLs (also expressed in ΔE units) increased with increasing frequency above around 2000 Hz. Thus the ratio between FB and FDL (FB/FDL) was fairly stable for frequencies up to 2000 Hz, and decreased at higher frequencies (similarly, Miller and Heise (1950) reported that the ratio ΔF/F at the trill threshold was nearly constant at 0.15 but decreased at higher frequencies). On average, the FBs below 2000 Hz were about 8 times larger than the FDLs. They concluded that the FB is not closely related to the FDL, and also that for low to medium frequencies it is not limited by discriminability. In a group of six hearing-impaired listeners, they found a much wider range of FB/FDL ratios than was seen in the normally hearing group; also, the ratios were not systematically related to the degree of hearing loss. They did find a weak positive correlation between FBs and FDLs, suggesting that enlarged FDLs (poor discriminability) may contribute to enlarged FBs; however, they concluded that their data provided only weak evidence of this.

In summary, auditory stream segregation in hearing impaired listeners has received some attention, but the precise relationship between hearing loss, frequency selectivity, and stream segregation is not yet entirely clear. The fact that the ratio between FB and FDL is almost always much greater than a value of 1 for both normal-hearing and hearing-impaired listeners (the FB is usually much larger than FDL) demonstrates that the ability to discriminate between two sounds is not sufficient *on its own* to enable stream segregation.

# Chapter 2: Cochlear implant listening

## 2.1 Cochlear implants: general principles and speech processing

Cochlear implants are now an established treatment for profound deafness, and all modern implant systems share common features, although minor design features vary between manufacturers. Sounds are picked up by microphones situated either behind the ear (on the top of a behind-the-ear speech processor) or in the ear canal (at the tip of an ear hook over the pinna, as in the case of the 'T-mic' used in the Advanced Bionics 'Harmony' speech processor). The speech processor applies analogue-to-digital conversion and a speech-processing strategy which determines how the input signal is represented in the pattern of stimulation applied to the array of electrodes in the cochlea. Crucially, the speech processor also applies compression and output limiting so that all electrical stimulation applied to each electrode is within the comfortable listening range for the implant user and does not exceed a maximum comfortable level. The output of the speech processor is transmitted via a radio-frequency carrier transcutaneously to the implanted receiver-stimulator package under the skin on the side of the head. The received signal is decoded and converted to pulsatile stimulation applied to the intracochlear electrodes. Design of the intracochlear electrode array varies somewhat between manufacturers; for example, the Cochlear Corporation Nucleus CI24 system (as used in all the experiments reported in this thesis) employs 22 electrodes in a 'modiolus hugging' array that attempts to place the electrodes close to the spiral ganglion that it is intended to stimulate.

A variety of approaches to speech processing strategies have been applied and performance with cochlear implant listening has improved steadily over the years since their invention. In all cases, the input signal is divided by band-pass filtering into channels that are allocated to individual electrodes. All represent the spectral content of sounds by place of stimulation in the cochlea. The number of channels used varies with device; the largest number of actual electrodes in clinical use is 22 as in the Nucleus CI24 system. Recently, there have been attempts to create up to 120 'virtual channels' using current steering in the Advanced Bionics device (more details below). In most cases, the stimulus delivered to the electrodes is pulsatile and charge-balanced biphasic pulses are used; loudness percepts can be varied through changes in pulse amplitude or pulse width, or a combination of both. In one commonly used strategy, 'ACE' (Advanced Combination Encoders) or '*n of m*', *n* channels are selected from a total possible *m* channels for stimulation in each 'sweep' across the electrode array (Wilson, 2006). Those channels with the highest envelope signals are picked out and stimulus pulses only delivered to a subset of the *m* channels available. In the Nucleus CI24 device, *m* is equal to 22 as there are 22 electrodes available for stimulation. The value of *n* in clinical use is variable between 6 and 16. Thus, spectral peaks in the stimulus waveform are represented in the pattern of stimulation across the electrode array, and this pattern is updated very rapidly with each sweep of stimulation from one end of the array to the other.

## 2.2 Pitch and music perception with cochlear implants

### 2.2.1 Pitch

Multi-electrode cochlear implants rely mainly on stimulation of discrete areas of auditory neurons and the tonotopic arrangement of the cochlea to convey a sense of pitch. However, the quality of pitch percepts that can be obtained by most implant users is variable and limited. The perception of pitch by cochlear implant users was explored extensively in the chapter by Moore and Carlyon (2005). As they point out, perception of pitch in normal (acoustic) hearing involves both timing (phase locking) and place cues. In cochlear implant listening, use of temporal coding is limited to low frequencies; most implant users are unable to detect changes in pulse rate above about 300 pulses per second (pps). Also, most commonly used speech processing strategies use the same pulse rate on all electrodes; pitch information based on phase-locking and temporal cues is not normally transmitted, and so implant listeners are dependent solely on place cues derived from peaks in the excitation pattern across the electrode array.

Nelson et al. (1995) reported very large variability in the performance of cochlear implant listeners in a pitch ranking task in which pairs of electrodes were stimulated sequentially and listeners responded according to which stimulus was higher in pitch or 'sharper'. They suggested that listeners who demonstrated steeper electrode ranking functions may have possessed a larger population of surviving neural elements, including functional cell bodies in the spiral ganglion that remain functionally organised tonotopically. Moore and Carlyon (2005) interpreted the data reported in the Nelson et al. (1995)

study to suggest that the threshold value for a median implant listener was about 1.2 mm along the basilar membrane, corresponding to a change in acoustic frequency of approximately 21%, based on Greenwood's (1990) equation. In contrast, the frequency difference limen for a normally hearing listener can be as low as 0.2%. Thus, the quality of percepts conveyed by place of electrical stimulation via a cochlear implant may only be loosely defined as pitch; although they can be described on a scale from 'low to high' or 'dull to sharp', it is not clear that they meet a strict definition of musical pitch (Moore and Carlyon, 2005). The distinction between the two dimensions of pitch 'height' and 'chroma' is important in this context. Implant listeners may perceive differences in timbral 'brightness' which may be similar to pitch height. In contrast, pitch chroma is more akin to musical pitch and may be inaccessible to them.

Another limitation imposed by the most commonly used speech processing strategies is that the frequency allocations to each channel are generally too wide to enable implant listeners to resolve individual harmonics; if two harmonics have frequencies that fall within the filter bandwidth allocated to a particular channel, they will be summed and produce identical stimulation on the electrode assigned to that channel. Also, spread of excitation in the cochlea around the stimulated electrode will lead to mixing of harmonics (this is discussed in more detail below). Furthermore, even when components of a complex tone are resolved, the log-spacing of channels in a cochlear implant does not allow a harmonic complex to be distinguished from any other kind. One consequence of this is the poor transmission of F0 by cochlear implants.

Some early speech-processing strategies (so called feature-extraction strategies) attempted to convey F0 via pulse rate; this approach has largely been abandoned in current approaches to speech-processing, in which pulse rate is normally constant across electrodes. Research efforts to improve the transmission of F0 for implant listeners are ongoing (e.g. Stickney et al., 2007). The inability of implant listeners to exploit F0 differences between competing voices is a major difficulty in multi-talker 'cocktail party' situations. Unlike normally hearing listeners, (e.g. Brokx and Nooteboom, 1982) there is some evidence that they cannot benefit from a gender difference when attempting to segregate speech sounds (e.g., Stickney et al., 2004). However, a recent report by Cullington and Zeng (2008) suggested that implant listeners *were* able to take advantage of a difference in F0 between target and masker talkers (a difference in mean F0 between around 100 Hz for a male voice and around 220 Hz for a female voice); when listening to male target speech, speech reception threshold was significantly better for a female than for a male masker.

In summary, perception of pitch by implant listeners is at best limited, and this severely restricts their ability to utilise important cues for sound segregation that are available in normal hearing. A great deal of research effort is concentrated on improving speech processing strategies so that F0 is conveyed better, but cochlear implants are not able to reproduce faithfully the pattern of auditory nerve activation that is elicited by sound in a normal ear; this represents a substantial limitation in the cochlear implant listener's ability to interpret complex sounds, and in particular to segregate competing sounds.

### 2.2.2 Music

Cochlear implant listeners are traditionally thought to obtain poor perception of music; indeed, the main thrust of implant development has focused on improved speech recognition, with appreciation of music as an almost accidental side-effect when it occurs, although recently more research effort has gone into improving the experience of music for implant users. Reports of poor music perception in implant listeners are several, and were reviewed by McDermott (2004). He summarised the findings of previous research in these terms:

a) On average, cochlear implant users perceive rhythm about as well as listeners with normal hearing; b) Even with sophisticated multiple-channel sound processors, recognition of melodies is poor, especially without rhythmic cues, with performance at little better than chance levels for many implant users; c) Perception of timbre, e.g. as used to identify musical instruments, is unsatisfactory; d) Implant users tend to rate the quality of musical sounds as less pleasant than do normal-hearing listeners; e) Auditory training programmes aimed at listening to music may help to improve enjoyment of music; f) Perception of pitch may be improved by enhancements to speech processing strategies; g) Combined electrical and acoustic hearing may help to improve music perception.

This disappointing aspect of cochlear implant performance is attributed to the inability of current implant speech processors and electrode arrays to provide listeners with sufficient spectral detail for music appreciation, combined with inadequate transmission of pitch, as already discussed. However, many implant users enjoy listening to music to varying degrees, although often their musical experience is restricted to genres that emphasise rhythm over complex harmonic structure. Also, as more research attention is directed to improving hearing for music in implant users, the importance of musical enjoyment is likely to receive greater emphasis. For example, Laneau et al. (2006) described a new sound processing scheme designed to optimise pitch perception and so hopefully perception of music. In their 'F0mod' scheme, slowly varying channel envelopes are modulated sinusoidally at the F0 of the input signal, with 100% modulation depth and in phase across channels in order to maximize temporal-envelope pitch cues. The intention was to provide better F0 information to the implant listeners that should aid hearing for musical pitch. They reported that melody recognition of familiar songs, with all rhythmic cues removed, was improved by using the F0mod scheme when compared to the 'ACE' (Advanced Combination Encoder) strategy, which provides no F0 information. Although this was a small study (6 implant listeners), the results suggest that gains in perception of music are possible with more sophisticated sound processing strategies, especially when they are designed to convey F0 information more explicitly.

## 2.3 Speech recognition in quiet and noise with cochlear implants

It is now well established that present-day cochlear implants can provide remarkably good speech recognition ability to many recipients in quiet environments. For example, Wilson and Dorman (2007) recently reported results on a set of nine tests of speech recognition from a Clarion CII implant user, compared with the results on the same tests carried out by a group of normally hearing listeners; they found that scores for this particular implant user were not different from the scores for the normal-hearing subjects for seven of the nine sub-tests. Differences were only apparent on two tests that involved sentences in a background of multi-talker babble, when the performance of the implant listener was impaired somewhat compared to the control group. However, in practice a wide variability in performance for implant listeners on tests of speech discrimination is common. Figure 2.1 shows the results on a standard measure of 'open-set' (sound alone, without lipreading) speech discrimination in quiet at 9 months post-implant surgery for a series of 125 adult users of the Nucleus CI24 implanted at the University Hospital Birmingham NHS Trust (Birmingham, UK). The median score in this series is 69% correct, but as can be seen the performance ranges from zero to 100% correct.

It is also well established that speech recognition performance with cochlear implants is often badly impaired by interfering noise. For example, Nelson and Jin (2002) showed that competing speech can impair performance in implant listeners even at a S/N ratio of +16 dB (more favourable than many real-life listening situations). Nascimento and Bevilacqua (2005) reported speech recognition performance in a series of 40 users of 5 different types of cochlear implant; their median sentence recognition scores fell from 90% in quiet to 50% correct in the +10 dB signal-to-noise condition, and only 20% correct in the +5 dB condition. They found no statistically valid difference in the performance with the different types of implant tested, suggesting that this problem is largely independent of the implant designs that are currently commercially available.

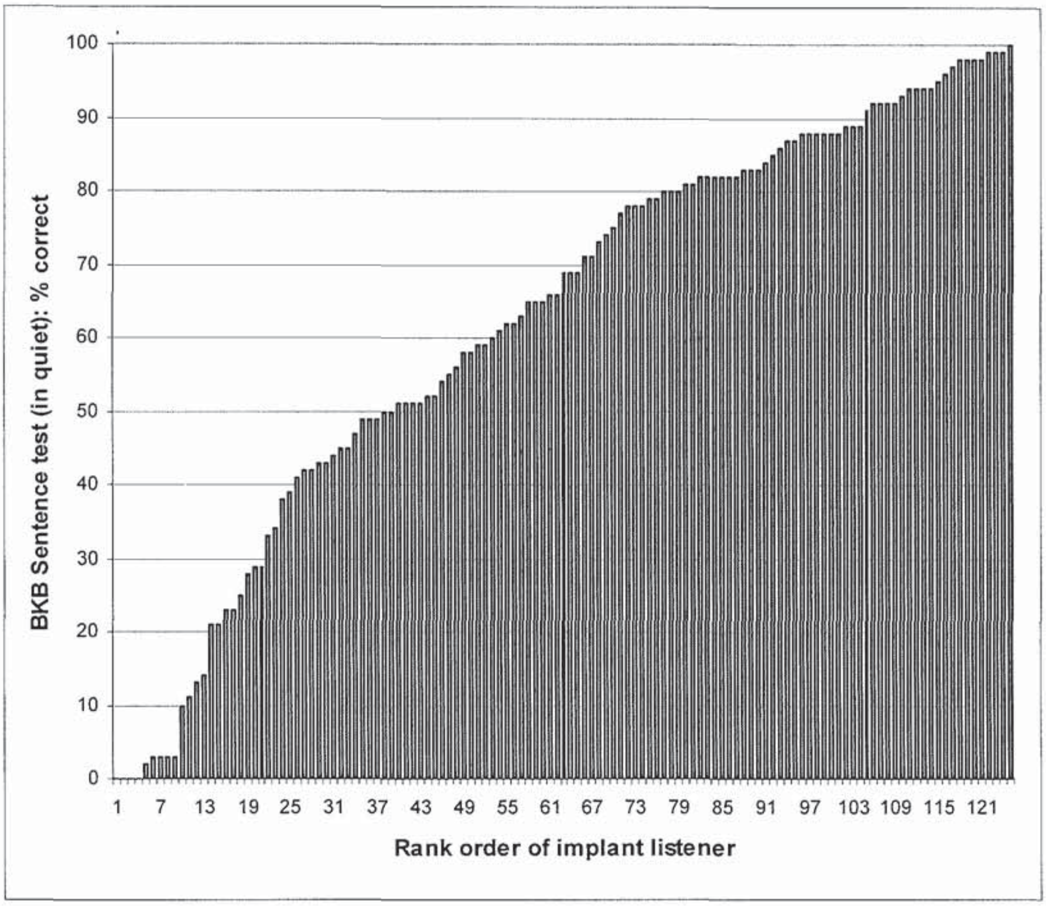


**Figure 2.1**. Open-set (sound alone) speech recognition scores for a series of 125 adult cochlear implant users with the Nucleus CI24 device, 9 months post surgery. Speech materials were BKB sentences, presented in the sound field at 70 dB(A). Implant users were able to adjust the sensitivity or volume settings on their speech processors as desired. The ordinate shows the % correct word score for a list of 30 sentences. Implant users are placed in rank order of their performance along the abscissa.

Many modern cochlear implant systems endeavour to improve the potential for speech understanding in noisy listening conditions by use of directional microphones, multi-microphone strategies or pre-processing of input sounds in the speech processor. For example, the Nucleus Freedom system gives the option of various strategies aimed at enhanced performance in noisy conditions. One option available is the directional 'Beam' strategy, which is a two-microphone adaptive beam-former that can provide some benefit in noisy conditions when listening to a single talker in front of the implant user. Spriet et al. (2007) reported improvements of up to 16 dB in the speech reception threshold (i.e. the signal-to-noise ratio at which a speech recognition score for words in sentences of 50% correct is obtained) for five Nucleus implant users when employing the 'Beam' strategy compared with the standard microphone.

Similarly, continuing research on development of speech-processing strategies is aimed at improving speech recognition performance in general, and particularly in noise. David et al. (2003) reviewed the performance of 139 cochlear implant users and concluded that improvements in performance could be attributed to evolving speech coding strategies and speech processors rather than to changes in patient selection criteria. One factor that is probably a major contributor to the impaired performance of cochlear implants in noise is the limited quality of the spectral information that implants are able to convey. Spectral resolution in implant listeners may be limited by: i) The number of functioning implanted electrodes (this varies between 16 in the Advanced Bionics System and 22 in the Nucleus Freedom system); ii) Quality and quantity of available neural tissue in the vicinity of the electrode array; iii) Current spread and spread of excitation between stimulating electrodes. Also, as noted earlier, the speech-processing strategies used by modern cochlear implants use input filters for individual channels that are typically too broad to resolve individual harmonics, and the spread of current across the cochlea would further increase the 'mixing' of harmonics (Moore and Carlyon, 2005).

Fu and Nogaki (2004) proposed that the susceptibility of implant listeners to noise may be attributable to their reduced spectral resolution, and the resulting spectral 'smearing' associated with channel interaction. They found that the performance of cochlear implant listeners in discriminating speech, in either steady state or gated noise, was comparable to that of normal-hearing listeners listening to an acoustic noise-band vocoder implant simulation, when spectral cues were 'smeared' by applying shallow filter slopes at -6dB/octave. Similarly, as previously mentioned, Gaudrain et al. (2007) reported two experiments with normally hearing listeners which suggested that reduction in spectral resolution may result in a reduced ability to form separate auditory streams, which they suggested may contribute to the difficulties of hearing-impaired and cochlear implant listeners hearing speech in noise. However, as already discussed, Rose and Moore (2005) did not find a clear relationship between fission boundaries and frequency difference limens in hearing-impaired listeners.

Shannon et al. (2004) pointed out that the number of spectral channels required for speech recognition depends on the difficulty of the situation; several studies have shown that for normal-hearing listeners, only around 4 to 6 spectral channels are necessary for 100% correct recognition of simple sentences in quiet (e.g., Shannon et al., 1995). The number of channels needed for speech recognition increases for more difficult speech materials, or for listening in noise; Shannon et al. (2004) suggested that up to 30 channels or more may be required for complex materials. Shannon (2005) also pointed out that speech and music have different requirements for spectral resolution; music requires at least 16 spectral channels even for identification of simple melodies. Dorman et al. (1998) reported the performance of a group of 21 normal-hearing listeners on a speech recognition task in which they were presented with sentences processed through a simulation of a cochlear implant speech processor. In quiet conditions, they achieved 100% correct recognition with only 6 channels. When listening to sentences in noise at a S/N ratio of -2 dB (not uncommon in everyday situations), their performance fell well below 100% correct recognition even with 20 channels available.

The actual number of independent channels available to cochlear implant listeners may be quite limited; Friesen at el (2001) reported, based on their study of 19 implant users with either the Nucleus 22 or Clarion devices, that even those implant listeners with the best speech recognition performance appeared to be unable to make use of more than seven to ten channels of spectral information, regardless of how many spectral channels were presented to them. They hypothesised that this may result partly from channel interactions. Baskent (2006) reported that speech recognition in the best cochlear implant listeners saturated around eight channels, and is not improved by the activation of more electrodes - presumably due to the reduced frequency selectivity resulting from channel interactions. Similarly, Wilson and Dorman (2007) suggest that the spread of stimulation of the auditory nerve produced in a multi-electrode implant array probably limits the number of perceptually separate channels to 4-8, even if more than eight electrodes are stimulated. It is therefore not surprising that cochlear implant listeners' ability to discriminate speech is so impaired in noisy conditions.

Despite this, efforts continue to increase the number of available spectral channels in implant devices; for example, Koch et al. (2007) reported results from 57 implanted ears for a new approach to current 'steering' recently developed by the Advanced Bionics corporation. Through simultaneous delivery of current to pairs of adjacent electrodes, it is hypothesized that the effective locus of stimulation can be 'steered' to sites between the real electrodes by varying the proportion of current delivered to each electrode of the pair. This can then provide an increased number of pitch percepts along the electrode array, theoretically much higher than that available from stimulation of the actual electrodes in the normal way. These are referred to as 'virtual channels'. Koch et al. found that the number of intermediate pitch percepts ranged from zero to 52, with an average number of between-electrode pitch percepts of 8.7. Overall this meant a range in the total number of pitches from 8 to 466. They suggest that 'the average cochlear implant user may have significantly more place-pitch capability than is exploited presently by cochlear implant systems'. However, evidence that an increased number of spectral 'channels' can effectively transmit more spectral information, and that implant listeners can make use of that information to obtain improved performance for speech (or music) recognition, is not yet available. In particular, it should be noted that increasing the number of discriminable pitches does not necessarily improve the spectral information available for more complex stimuli at all.

Summary:

i) Speech recognition performance with cochlear implants has improved significantly with advances in speech coding strategies;

ii) Despite this, speech recognition in implant listeners is generally badly impaired by background noise;

iii) Evidence suggests that the number of perceptually distinct spectral channels available to implant listeners is probably limited to around eight, which is not sufficient for reliable speech recognition in noise (depending on the signal-to-noise ratio);

iv) Spread of excitation in the cochlea leads to channel interactions and spectral 'smearing' which severely impair the ability to discriminate speech in noise.

v) Cochlear implant listeners are unable to make use of differences in F0 as a means of segregating concurrent speech.

## 2.4 Auditory stream segregation in cochlear implant listeners

Cochlear implant listeners clearly need to deal with the same listening conditions as people with normal hearing; the ability of the auditory system to assign sounds perceptually to different sources is relevant and important, irrespective of whether the listener is hearing via electrical stimulation or through natural means. As the vast majority of implant users have only one implant, and are therefore listening monaurally, they do not have access to any binaural cues that might aid perceptual segregation of concurrent or sequential sounds.

Few studies of stream segregation in cochlear implant listeners have been reported. Probably the first example is the study of Chatterjee and Galvin (2002). They used repeating patterns of loudness-matched stimuli composed of two or three different tones (pulse trains) and varied the tonotopic distance (electrode pair separation) between them. Before each test sequence, listeners heard a "preview" sequence with a rhythm corresponding to a subset of the tones and they were asked to judge whether or not they could hear this rhythm within the test sequence. A high proportion of positive responses was taken as evidence that the particular subset of elements could be heard out as a separate perceptual stream. Chatterjee and Galvin reported the results for a TRT of 100 ms (tone duration = 50 ms, inter-tone interval = 50 ms) and found that the proportion of positive responses increased with tonotopic distance, as would be expected for normal-hearing listeners. Ratings of the degree of perceived segregation showed a similar dependence on tonotopic distance. On the basis of their results, Chatterjee and Galvin (2002) suggested that sequential stream segregation was achievable in cochlear implant listeners. However, their experimental design did not allow for a definitive conclusion that *automatic* stream segregation was taking place; i.e., that the test sequence breaks into two perceptual streams that cannot voluntarily be recombined by the listener (Bregman, 1990). First, they did not report the effects of varying the rate of presentation for their sequences of stimuli. This is important because, as noted earlier, the temporal coherence boundary is known to be strongly dependent on the TRT (van Noorden, 1975). Second, their task involved the selection of a subset of elements from a larger sequence. It is known from studies of normal-hearing listeners that it is possible to focus attention on a subset of tones in a sequence when they differ in frequency by only a few semitones (e.g., Cusack and Roberts, 2000).

More recently, Chatterjee, et al. (2006) reported the results of two stream segregation tasks in five cochlear implant listeners, using a research interface that allowed careful control of the stimuli applied to individual electrodes. In their first experiment, they used an ABA-ABA... presentation sequence of alternating stimuli, at a fixed rate (duration of individual stimuli was 50 ms, with 50 ms inter-stimulus intervals, equivalent to a TRT of 100 ms). This fixed rate was apparently chosen because the authors had found that it produced strong segregation when they listened to acoustic stimuli at the same rate. Stimulus A always corresponded to bipolar stimulation of an electrode in the centre of the 22–electrode array (electrode 10). Stimulus B was then varied across a range of electrodes either more apical or more basal to A. The listeners heard varying durations of the ABA-ABA... sequence (between 1.2 and 3.9 seconds), at the end of which they indicated whether they had heard two separate streams. The overall percentage of times (across multiple presentations) that each sequence was reported as two streams was then recorded; one yes/no response to the question 'Did you hear two separate streams' was elicited after each sequence presentation. In a separate task, the same listeners were asked to report if they could hear any change in sound (e.g. pitch, timbre) when a single ABA triplet for each electrode combination was presented: this gave a discrimination measure derived from the percentage of 'yes' responses obtained from a minimum of 20 presentations. Finally, they listened to the stimuli A and B in isolation and rated the perceptual magnitude of the difference between them on a scale from 0 to 100 (presumably a score of 0 represented identical sounds and 100 represented totally different sounds). The results show, for most of the listeners, a dependence of the reported stream segregation on electrode pair separation; the percentage of reported segregation tended to increase with increased separation of A and B across the electrode array. However, there was substantial inter-listener variability in their results. One of these listeners showed evidence of a build-up of streaming, by reporting two streams much more often for the two longest sequence durations (2.7 and 3.9 seconds) than for the shorter sequences. Chatterjee et al. suggest that this was evidence of stream segregation. However, the other four listeners showed no effect of sequence duration on reported stream segregation; this implies that the task was not measuring stream segregation.

In their second experiment, in which only one listener took part (the same listener who had shown an effect of sequence duration in the 1st experiment), stimuli A and B were presented to the same electrode pair, but instead were amplitude-modulated (AM); for stimulus A, the AM rate was fixed at 100 Hz, while the AM rate for stimulus B was varied between 20 and 160 Hz. This difference in temporal envelope between A and B provided the listener with a perceptually salient difference between A and B. As before, the listener reported when they could hear two distinct streams at the end of sequences of varying duration (0.9 seconds to 3.9 seconds). The results showed a clear dependence of the percentage of stream segregation reported on the difference in AM rate between A and B. Specifically, the proportion of stream-segregation reports rose steeply as the AM rate for stimulus B was varied outside the range +/- 20 Hz around the 100 Hz value used for A; this effect was observed at all the sequence durations above 1 second, and as before for this listener the most segregation was reported for the two longest sequences (2.7 and 3.9 seconds). The results of the discrimination measure (in which the listener heard one ABA triplet from the sequence and reported if s/he could hear a change in the sounds within the triplet) showed a very similar pattern; the proportion of 'yes' responses also rose steeply when the AM rate for stimulus B was greater than 120 Hz or less than 80 Hz.

Chatterjee et al. suggested that these results showed that it is possible for implant listeners to perceptually segregate sequential stimuli based on differences in either cochlear place or temporal envelope; however, they state, correctly, that it is not possible to be sure if their listeners actually experienced streaming. They admitted that if the listeners did not know what to listen for, they might have resorted to listening for pitch differences, rather than stream segregation, when performing the task. Although their results were consistent with auditory stream segregation, they did not provide conclusive evidence of it. As mentioned above in relation to the Chatterjee and Galvin (2002) study, this study also did not measure the effect of rate of presentation (TRT) on reported segregation, or any of the other properties of streaming that were discussed above.

## 2.5 Objective measures of stream segregation in cochlear implant listeners

Hong and Turner (2006) reported the results of a temporal discrimination task designed to provide a measure of stream segregation in cochlear implant listeners. The rationale for the use of such a task is that perceptual properties of auditory events are computed within but not across streams. Therefore, stream segregation should lead to worse performance on a task requiring judgements of the relative timing of sounds (e.g., Warren et al., 1969; Bregman and Campbell, 1971). Hong and Turner's first experiment was based on the task introduced by Roberts et al. (2002), in which listeners hear in succession on each trial two sequences of rapidly alternating tones (A and B). The rhythm remains isochronous throughout in one sequence; in the other it begins as isochronous but becomes progressively irregular after several AB cycles. The extent of the delay applied to sound B was varied using an adaptive staircase and the task of the listener was to identify the irregular interval. By introducing the delay on sound B only after several AB cycles, sufficient time was allowed for the strength of stream segregation to build up before the two sequences began to differ (Bregman, 1978; Anstis and Saida, 1985).

Hong and Turner used acoustic stimuli that were presented via loudspeaker, and their cochlear implant listeners used their normal speech processors and speech-processing strategies. Hong and Turner reported a strong and progressive increase in temporal discrimination thresholds as the frequency separation between tones A and B was increased, which is consistent with the idea that stream segregation had occurred in their implant listeners. However, this finding is not conclusive in itself, because thresholds for detecting a temporal gap between two isolated pure tones rise for normal-hearing listeners as the frequency separation is increased (see, e.g., Grose et al., 2001). Similarly, gap detection thresholds rise as the physical separation of stimulated electrode pairs is increased in cochlear implant listeners (Hanekom and Shannon, 1998). Therefore, in their second experiment, Hong and Turner measured the extent to which detection of the delay on tone B was dependent on frequency separation in the context of an isolated ABA triplet, for which there was insufficient time for a significant build-up in the tendency for stream segregation. The dependence they observed was shallower than in the first experiment, which supports their conclusion that the cochlear implant listeners in their first experiment were experiencing automatic stream segregation. However, in their second experiment only three listeners took part in the short-sequence task that is key to sustaining the conclusion that stream segregation influenced judgements of the longer sequences. Another weakness of Hong and Turner's study is that their implant listeners were presented with acoustic stimuli which would then have been subject to the speech-processing strategies programmed into their implants. This may have had the advantage of simulating real-world listening, but had the clear disadvantage that control of the actual stimuli delivered to their implanted electrodes was difficult and unpredictable, due to the variable and unknown effects of the speech processing strategy in use. For this reason, studies of stream segregation in implant listeners should preferably employ careful manipulation of the stimuli applied to individual electrodes.

Summary:

i) There have been some preliminary reports of experiments with results that are *consistent* with auditory stream segregation taking place in cochlear implant listeners;

ii) Despite this, the results published to date do not provide convincing evidence of automatic stream segregation in the majority of implant listeners;

iii) Further evidence drawn from a variety of experimental approaches is required to fully evaluate the extent to which implant listeners are able to make use of the cues to segregation that have been measured and reported for normally hearing listeners.

## 2.6 Measures of channel interaction in cochlear implants

### 2.6.1 Psychophysical approaches

As already discussed, channel interactions and spread of excitation between/across intra-cochlear electrodes are thought to reduce the spectral information that is conveyed and so to be critical to the performance of implant listeners in natural, noisy environments. Despite attempts to design electrode arrays that reduce current spread such as the pre-curved, 'modiolus-hugging' Nucleus 'Contour' or Advanced Bionics 'Helix', in practice overlapping populations of nerve fibres are inevitably stimulated. A number of approaches have been employed to evaluate the extent of the resulting channel interaction. For example, Hanekom and Shannon (1998) used gap detection thresholds as an indirect measure of electrode interaction. Their hypothesis was that gap detection thresholds would be shortest when the two stimuli that bound the gap stimulate the same electrode; if the two stimuli before and after the gap are presented to electrodes that are spatially separated, the amount of overlap in the neural populations stimulated by each will be reduced, the two stimuli will sound dissimilar, and this will lead to increased gap detection thresholds. The result is a dependence of the gap thresholds on electrode separation, providing a form of 'psychophysical tuning curve'. They reported considerable variability in the gap detection curves for the three listeners who participated in their study, and suggested that the better implant users demonstrated 'sharper' tuning (indicative of a smaller area of neural activation). However, more data from a larger group of implant listeners would be required before a definitive conclusion could be drawn.

Another approach to estimation of channel interaction is forward-masked excitation patterns, as described by Chatterjee and Shannon (1998). In this method, a masker stimulus (300 ms duration) is applied to one electrode prior to a probe stimulus (20 ms duration) on either the same electrode pair or any of the other available electrodes across the implant array. The resulting increase in threshold for detection of the probe produced by the presence of the masker can then be measured and plotted as a function of the location of the probe electrode. The largest threshold increase can be expected when masker and probe stimulate the same set of neurons (or channel), and smaller increases when masker and probe activate different neural populations or channels. Chatterjee and Shannon reported the results for 4 implant listeners, which again demonstrated inter-listener variability; for all the participants a peak in threshold increase was found with probe and masker on the same electrode pair, with a sloping function either side towards more apical or basal probe electrodes. They suggested that the observed variance in shape of these curves between listeners was related to differences in aetiology, nerve survival, and electrode insertion depth. However, they did not claim to demonstrate a clear relationship between these psychophysical measures and speech recognition performance.

More recently, Stickney et al. (2006) investigated the relationship between measured channel interaction and various measures of speech recognition in eight users of the Advanced Bionics 'Clarion' cochlear implant. They quantified electrical-field interaction by measurement of the detection threshold for a probe stimulus applied to one electrode, simultaneous with a sub-threshold 'perturbation signal' applied simultaneously to another electrode in the array. They then calculated an 'Interaction Index' from the difference in threshold obtained with an in-phase and out-of-phase perturbation signal. They reported significant negative correlations between the mean interaction index and discrimination performance for vowels, consonants and sentences (r=-0.74, -0.83 and -0.68 respectively) in those listeners employing the 'SAS' (Simultaneous Analogue Stimulation) speech processing strategy, but found no such correlation in listeners using the CIS (Continuous Interleaved Sampling) strategy. They suggested that one explanation for this finding is that the CIS strategy avoids electrical-field summation as it uses sequential rather than simultaneous stimulation pulses. However, the numbers of listeners in each group are small, so it would be difficult to draw clear conclusions from their data.

### 2.6.2 Electrophysiological approaches

An electrophysiological approach to estimation of channel interaction was described by Cohen et al. (2003). The method they introduced was based on the 'subtraction' paradigm implemented in the Neural Response Telemetry ™ (NRT ™) system for recording the electrically evoked compound action potential (ECAP) in the Nucleus CI24 cochlear implant. NRT recording involves presentation of a stimulus to one electrode in the intra-cochlear array and recording of the ECAP generated from another, usually neighbouring or spatially close electrode. The subtraction method reduces the large stimulus artefact, allowing recording of the neural response alone. In this method, three main responses are obtained from stimulation on one of the intra-cochlear electrodes, as follows:

A= Probe pulse;
B= Probe pulse preceded by a masker pulse;
C= Masker pulse alone.

The probe alone stimulus (A) produces a response including both stimulus artefact and the neural response. In the B condition, the masker pulse causes most of the nerve fibres in the vicinity of the stimulating electrode to be in a refractory state, so that they do not respond to the probe pulse that follows, if the masker-probe gap is sufficiently brief. The neural response on its own is then obtained in the recording software by subtraction, according to the formula A-(B-C). The amplitude of this neural response can then be measured, and measurements at a series of stimulus levels can be used to generate an amplitude growth function.

In the subtraction paradigm, maximum amplitude neural responses (ECAPs) are recorded when probe and masker are on the same electrode (thus stimulating the same neural population), the masker is at the same level as the probe, and there is a short masker-probe interval (e.g. 500 μs). If the electrode to which the masker pulse is presented is widely physically separated from the probe electrode, the neural populations stimulated by each do not overlap, resulting in an absent ECAP response. If the masker and probe electrodes are different, but remain relatively close along the electrode array, the neural populations they stimulate will partially overlap, resulting in an ECAP response that is recordable but reduced in amplitude compared with the maximum response. Thus, by recording the ECAP amplitudes for a range of masker locations along the array while maintaining a fixed probe electrode, the spread of excitation in the cochlea can be estimated. The amplitude of the measured ECAP response is an indirect measure of the overlap of the neural excitation produced by the probe and masker electrodes. Cohen et al. referred to this as the 'Variable masker, fixed recording location' or VMFR spatial measure, and reported good agreement between the excitation profiles derived using it and those obtained from psychophysical forward masking in seven implant users. They also demonstrated an increase in the width of the ECAP profile with higher MCLs (Most Comfortable Levels of stimulation) for individual listeners, and suggested that this was associated with poorer neural survival. This method of estimating channel interaction has advantages over the psychophysical forward–masking method described above as it is quicker, and more reliable and objective.

The question of whether estimates of channel interaction measured using the 'VMFR' method described by Cohen et al. (2003) have any relationship with pitch-ranking or speech recognition performance was investigated by Hughes and Abbas (2006). They hypothesised that significant channel interaction, associated with broad ECAP excitation profiles and significant overlap between stimulated populations, might result in spectral smearing, reduced electrode pitch discriminability, and poorer speech perception. To test this, they measured the ECAP amplitudes with variable masker electrode positions (as described by Cohen et al., 2003) in ten adult Nucleus CI24 implant users. They then calculated the width of the ECAP amplitude function at a point equal to 75% of the maximum normalised amplitude, expressed in terms of numbers of electrodes. The results of the ECAP profiles measure were then compared with pitch-ranking ability measured using a two-interval, two-alternative electrode-ranking task in the same group of listeners. The raw results on this second task were the percentage of responses where each of the other randomly presented electrodes across the array were judged as higher in pitch than a standard electrode. These were converted into z-scores and the slopes were calculated for a linear regression fitted to plots of z-score vs. electrode number. The hypothesis that greater channel interaction (leading to wider ECAP excitation profiles) should be associated with poorer electrode pitch discrimination would predict a significant positive correlation between the width of the ECAP profiles and the z-score slopes derived from the electrode-ranking task. However, none of the ten listeners in this study produced this result. In fact, although three listeners showed a moderate positive correlation between ECAP profile width and pitch discrimination, another three showed a similar (or in two cases greater) negative correlation between the two measures. There were also no significant correlations between the pitch-ranking performance and three separate measures of speech recognition ability, and no significant correlations between ECAP profile width and the same speech recognition measures.

A number of explanations for the lack of support for the predictions made were discussed by the authors. These relate to the nature of the electrode-ranking task used, which specifically asked the implant listeners to rank electrodes on the basis of pitch, as opposed to a same/different judgement that had been used in other studies (where listeners may have based their judgements on other perceptual differences). Also, they suggested that while the ECAP profile method is an entirely peripheral measure, electrode pitch ranking involves more central processing. The lack of a relationship between the ECAP profile widths and speech recognition in quiet is perhaps less surprising than the finding for pitch ranking. Speech discrimination performance in cochlear implant listeners is known to be influenced by a wide range of variables, each of which individually contributes only a small amount to the variance in performance. It may be that speech measures that are more dependent on transmission of spectral information, such as vowel identification, would be more appropriate for correlation with the ECAP data than standard speech materials.

## 2.7 Overview of thesis aims, methodology and key findings

As already discussed, cochlear implant listeners receive an impoverished auditory input compared with normal hearing, with limited spectral resolution and a restricted number of independent channels. Although many implant listeners can obtain good speech recognition performance with their devices in quiet listening conditions, they often experience particular problems listening to speech in noisy situations, and poor perception of music. There is also relatively little convincing evidence to date of automatic auditory stream segregation in implant listeners. A number of cues which contribute to stream segregation have been shown to be used by normal-hearing listeners, and experimental approaches to measuring them either directly or indirectly are well established. These can be adapted to investigate segregation in cochlear implant listeners.

The broad aims of this thesis are:

a) To explore the extent to which cochlear implant listeners are able to access any of the cues to auditory stream segregation that are known to be available to listeners with normal hearing, using a variety of experimental approaches;
b) To evaluate the properties of sounds that enable cochlear implant listeners to perform auditory scene analysis;
c) To investigate whether measures of performance on tasks involving stream segregation relate to other performance measures in cochlear implant listeners, e.g. speech recognition;

d) To investigate the degree of channel interaction caused by spread of excitation in the cochlea in cochlear implant stimulation using electrophysiological measures.

- *Experiment 1* (chapter 3) employed the often-used paradigm of repeating ABA..ABA... sequences of non-overlapping pure tones (c.f. van Noorden, 1975; Singh and Bregman, 1997). This provided an evaluation of whether there is any evidence of automatic, sequential stream segregation in implant listeners in a task in which the stimuli are relatively simple and easily defined. The individual channels in a multiple-electrode cochlear implant are each allocated a specific frequency band; because of this, the pure tones used were 'targeted' at the centre frequencies of the particular electrodes that were intended for stimulation. This method was also used for experiments 2, 3 and 4. Thus, electrode separation was used in place of frequency separation. Also, a standard pulse rate (900 Hz) was used throughout all the experiments described here; in some cases, this did not coincide with the pulse rate normally used by the implant listeners in normal daily use. No evidence was found that any of the listeners were disadvantaged by this approach, and it was considered more important to standardise this variable rather than to use listeners' normal pulse rates, which would have inevitably introduced another variable into the experimental design. In experiment 1, the effect of presentation rate on reported segregation was explored, and the incidence of perceptual ambiguity was measured. Little evidence of an effect of rate or of perceptual ambiguity was apparent. Thus, strong evidence that stream segregation was not generally affecting performance on this task was found.

- *Experiment 2* (chapter 4) employed a pitch-discrimination task. If the data obtained in experiment 1 provided a measure of channel discrimination rather than primitive stream segregation, then overall performance on a more straightforward pitch task, and also individual differences in performance, should correlate strongly with the corresponding reported segregation scores obtained in experiment 1. The task was in effect an electrode discrimination task; listeners reported differences in pitch (or sharpness/brightness- as discussed above, cochlear implant listeners may perceive differences between stimulation on different electrodes in terms of timbre rather than musical pitch) between stimuli presented on two electrodes sequentially. They also rated their confidence in their pitch rankings. Although insertion of an implant electrode array does not guarantee a monotonic relationship between electrode number and pitch, it is assumed here that more basal electrodes (lower numbers) evoke higher pitches than more apical electrodes (higher numbers). A positive relationship between the confidence in pitch ranking judgements and reported segregation scores from experiment 1 was found. The results of experiment 2 are consistent with the view that the results of experiment 1 primarily reflected the channel discrimination abilities of these listeners, rather than the effects of automatic stream segregation.

- *Experiment 3* (chapter 5) employed a temporal discrimination task in which stream segregation would be expected to be detrimental to performance (c.f. Vliegen et al. 1999, Roberts et al., 2002). This provides an indirect measure of primitive stream segregation, in which segregation results in impaired performance despite the efforts of the listener. The task required the detection of a temporal shift (delay) in the onset of tone B in an alternating tone sequence ABA..ABA..ABA. If stream segregation was occurring for the implant listeners, increased electrode separation between A and B should lead to reduced performance (i.e. greater threshold delays), and this effect should be greater at the end of a sequence lasting a few seconds (i.e., long enough for segregation strength to build up) than for a brief ABA triplet. This was not found; although a clear dependence of threshold delays on electrodes separation was seen, there was no significant effect of sequence duration. Thus, this experiment failed to provide any evidence that automatic stream segregation was taking place.

- *Experiment 4* (chapter 6) used a melody discrimination task similar to that described by Dowling (1973) and Cusack and Roberts (2000). In this case, segregation of the target tones in a simple melody from distractor tones temporally interleaved with them should improve performance. This was a selection task, which measured the ability of implant listeners to ignore the distractors and discriminate between two simple target melodies. Distractors were either on the same range of electrodes as the melodies (same pitch range), or on spatially distinct electrodes (in either a higher or lower pitch range than the target tones). If implant listeners were able to use differences in pitch to segregate a target melody from distractors, their performance should be significantly better when the distractors were on a different set of electrodes to the melodies. This was not found; listeners showed poor performance even when the distractors were spatially remote (i.e. assumed to have different pitches) from the target stimuli, except when the distractors were substantially attenuated.

- *Experiment 5* (chapter 7) employed a simultaneous grouping task, which concerned the ability of implant listeners to use abrupt changes in level to pick out components from a complex tone. In normally hearing listeners, sudden increases in the amplitude of pure-tone components allow them to be perceptually isolated from an on-going complex tone, and more rapid increments (shorter rise times) enhance this effect (Bregman et al., 1994b). This has been described as a 'sudden change' segregation effect which leads to a 'resetting' of the pitch-analysis mechanism. This experiment investigated whether similar effects can be demonstrated in cochlear implant listeners. For this experiment, stimuli were created with the 'Impress' software and presented via a 'Spear' research speech processor, which allowed precise control of all stimulus parameters on each electrode. Some listeners were able to benefit from sudden-onset increments in level, but overall no significant effect of attack time was found. Most of the listeners were unable to perform this task at all when decrements in level were used.

- *Experiment 6* (Chapter 8) reports the results of objective, electrophysiological measures of spread of excitation using the 'Variable masker, fixed recording location' method for recording the ECAP response with Neural Response Telemetry $^{TM}$ (NRT) in a group of cochlear implant listeners. The width of the NRT profile measured using this method provides an estimate of the spread of excitation from individual stimulated electrodes, and so an indirect estimate of channel interaction. Wider profiles, indicating greater spread of excitation and channel interaction, might be associated with poorer electrode discrimination, and so poorer performance on any tasks that depend on good discriminability between electrodes and spectral resolution, including stream segregation. The results of the ECAP profile widths are compared with data from the other experiments related to reported segregation, electrode pitch ranking, and speech recognition performance with the implant. No significant relationship between ECAP profile widths and speech recognition ability with the implant was found. However, some correlation was found between the slopes of the ECAP profiles and reported segregation and confidence in electrode pitch ranking judgements.

# Chapter 3: Experiment 1

## 3.1 Introduction

The simplest form of perceptual grouping that can be explored in implant listeners is the perceptual organisation of sequences of pure tones. These stimuli have the advantage of allowing an exploration of the perceptual effects of reduced signal quality whilst avoiding the complication of temporal overlap between individual sounds. This approach has a long track record in studies of stream segregation in normally hearing listeners (e.g. van Noorden, 1975; Bregman, 1978; Anstis and Saida, 1985; Singh and Bregman, 1997; Vliegen and Oxenham,1999; Carlyon et al., 2001). It was also used by Chatterjee et al. (2006) to try and evaluate the extent of stream segregation in cochlear implant listeners, as discussed in chapter 2. They reported the results of two stream segregation tasks in five cochlear implant listeners (see chapter 2 for more details). They suggested that their results showed that it is possible for implant listeners to perceptually segregate sequential stimuli based on differences in either cochlear place or temporal envelope; however, they stated, correctly, that it is not possible to be sure if their listeners actually experienced streaming.

## 3.2 Overview of experiment 1

The aim of this experiment was firstly to provide, if possible, more definitive evidence of automatic stream segregation in cochlear implant listeners. In order to provide convincing evidence of auditory stream segregation, three of the well-documented factors that were known to influence this phenomenon from studies using normal-hearing listeners were manipulated:

1. The effect of frequency separation. In normal hearing, increased frequency separation increases the probability of two distinct auditory streams being heard (e.g., Bregman and Campbell, 1971). In implant listeners, increased electrode separation would be expected to give rise to more stream segregation (with the caveats discussed in more detail below).

2. The effect of rate of presentation on the probability of stream segregation. The temporal coherence boundary (van Noorden, 1975) has been shown in normal-hearing listeners to increase with increasing TRT (i.e., reduced rate of presentation). This means that, within a certain range, TRT should affect the probability of alternating tones being reported as two streams if primitive stream segregation is taking place.

3. For frequency separations lying between the temporal coherence boundary and the fission boundary, the percept heard should be ambiguous (e.g., Anstis and Saida, 1985), such that it can 'flip' randomly between integration (one stream) and segregation (two streams). In a task in which the listener is required to respond every time they perceive a change in organisation from one to the other, this should typically lead to several key-presses over the course of 30 seconds.

## 3.3 Method

In this experiment, sequences of temporally non-overlapping pure tones were employed, thus enabling the investigation of sequential grouping. Similar stimuli to those described by van Noorden (1975) were used, i.e. a repeating ABA-ABA- sequence of pure tones, where the frequency of tone A remained constant throughout the experiment but that of tone B was varied.

### 3.3.1 Stimuli

Characteristics of the stimuli are summarized in Tables 3.1 and 3.2. All stimuli used were pure tones at the frequencies corresponding to the centre frequencies of the frequency allocations for each of the 22 channels programmed into the speech processor 'Map' used throughout experiments 1 and 2. Although spectral 'splatter' theoretically should not be a risk for stimuli presented via electrical stimulation, 10-ms rise and fall times were used for each tone. In order to investigate the effect of rate of presentation on reported stream segregation, four different TRTs were used: 50 ms, 100 ms, 150 ms, and 200 ms (see Table 3.1). These values exceed slightly the range of TRTs used by van Noorden (1975), which was 60 to 150 ms. He showed an approximately two-fold increase in the temporal coherence boundary (from around 6 to 12 semitones) for a change in TRT from 100 to 150 ms, thus demonstrating a strong effect of rate on primitive stream segregation in normal-hearing listeners.

To ensure that different stimuli resulted in stimulation on different channels in experiments 1 to 3, their frequencies were selected to correspond to the centre frequency of each channel in the experimental speech processor used (see Table 3.2).

**Table 3.1**: Temporal characteristics of the stimuli.

| TRT (ms) | Onset ramp (ms) | Plateau (ms) | Offset ramp (ms) | Within-triplet silent intervals (ms) | Between-triplet silent intervals (ms) |
|---|---|---|---|---|---|
| 50 | 10 | 20 | 10 | 10 | 60 |
| 100 | 10 | 40 | 10 | 40 | 140 |
| 150 | 10 | 70 | 10 | 60 | 210 |
| 200 | 10 | 100 | 10 | 80 | 280 |

**Table 3.2**: Frequency characteristics of the stimuli and their relation to the implant channels.

| **Channel/electrode number** | **Lower frequency boundary (Hz)** | **Upper frequency boundary (Hz)** | **Channel centre frequency and frequency of the pure tone stimuli (Hz)** |
|---|---|---|---|
| 1 (most basal) | 6938 | 7938 | 7438 |
| 2 | 6063 | 6938 | 6500 |
| 3 | 5313 | 6063 | 5688 |
| 4 | 4688 | 5313 | 5000 |
| 5 | 4063 | 4688 | 4375 |
| 6 | 3563 | 4063 | 3813 |
| 7 | 3063 | 3563 | 3313 |
| 8 | 2688 | 3063 | 2875 |
| 9 | 2313 | 2688 | 2500 |
| 10 | 2063 | 2313 | 2188 |
| 11 | 1813 | 2063 | 1938 |
| 12 | 1563 | 1813 | 1688 |
| 13 | 1313 | 1563 | 1438 |
| 14 | 1188 | 1313 | 1250 |
| 15 | 1063 | 1188 | 1125 |
| 16 | 938 | 1063 | 1000 |
| 17 | 813 | 938 | 875 |
| 18 | 688 | 813 | 750 |
| 19 | 563 | 688 | 625 |
| 20 | 438 | 563 | 500 |
| 21 | 313 | 438 | 375 |
| 22 (most apical) | 188 | 313 | 250 |

The speech processor was programmed so that it would deliver stimulation on only one channel at a time (in normal use 6 electrodes or more are activated, selected from the full set of available electrodes). The 'Map' used was programmed with the number of 'maxima' set at a value of 1 so that only one electrode could be stimulated at a time, using the 'Advanced Combination Encoder' (ACE) strategy and Nucleus R126 programming software. For all listeners, the pulse rate used in this study was 900 pulses per second and the pulse width used was 25 μs per phase. In order to verify that the selected channels were indeed activated as intended by the pure tone stimuli described, all experimental stimuli were routed through an experimental speech processor and the output was analysed using a 'dummy' cochlear implant within the manufacturer's computer interface, which allows a stimulus frame-by-frame listing of the output of the implant transmitter coil and the generation of an 'Electrodogram' which illustrates visually the output on each electrode over a selected time window (see Figure 3.1).

### 3.3.2 Loudness balancing

Prior to the experiment, comfortable and equal loudness levels were measured for each listener for each of the pure tones using the implant manufacturer's programming software. Individual electrodes were set at the maximum comfortable loudness level or 'C level' (the highest stimulus level that remained comfortable). This was achieved by a 'sweep' across channels, as many times as required to ensure equal and comfortable loudness for all the stimuli. In this procedure, 500 ms duration stimuli, with 500 ms inter-stimulus gaps, were applied sequentially to each electrode in turn, starting at the apical end of the electrode array and progressing to the basal end. The listeners reported the loudness of the stimulation on each electrode, and adjustments were made to individual levels as necessary to provide equal loudness across the entire array. All stimuli were then presented at these levels for the rest of the experiment. Stimuli were delivered via a computer-controlled sound card (16-bit resolution, 20-kHz sampling frequency) to the external input socket of the speech processor, via an electrically isolated adaptor cable (as supplied by the manufacturer for use with the Nucleus implant system). The control knob on the adaptor cable and the sensitivity control on the experimental speech processor were set to maximum, thus ensuring that stimulation always occurred at the appropriate C levels.

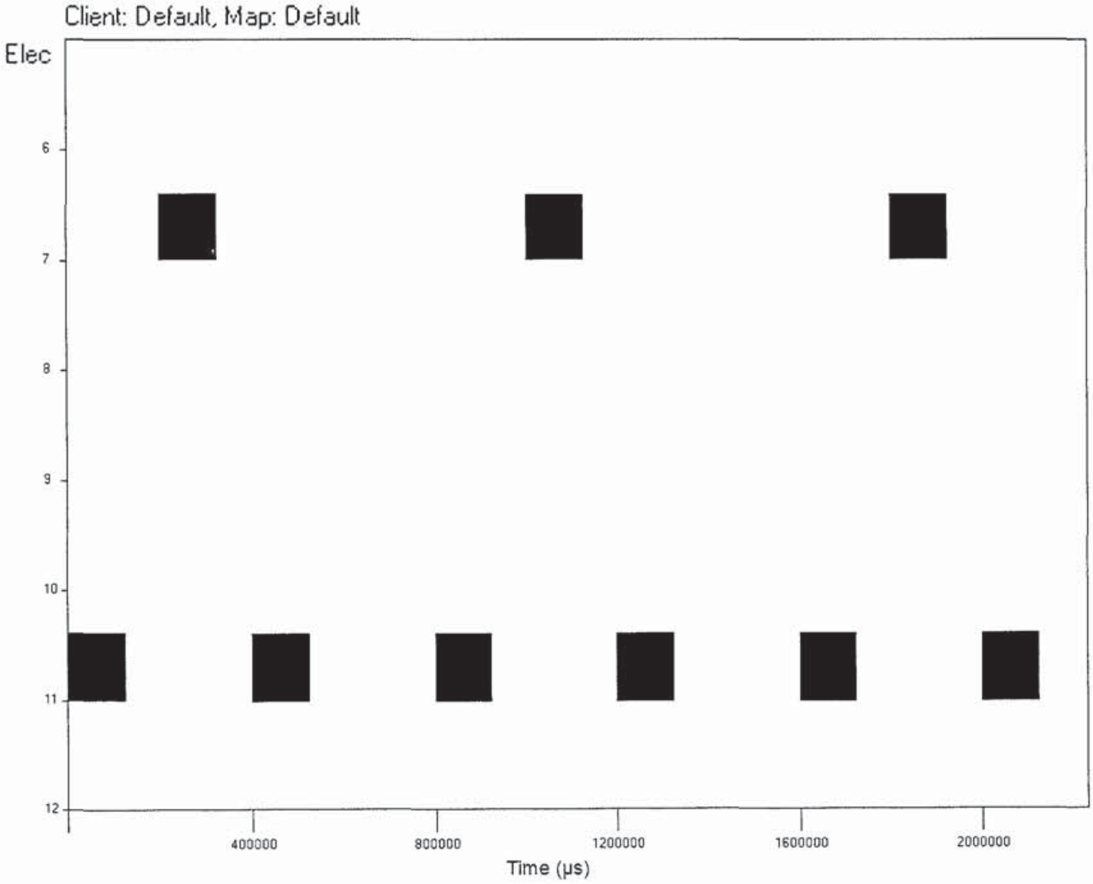


**Figure 3.1.** An example "electrodogram" displaying the output of the experimental speech processor and transmitter coil for a time window of approximately 2.2 seconds. The output was taken from within an experimental run comprising ABA triplets with a TRT of 200 ms. The ordinate shows electrode number and the abscissa shows time elapsed in μs. Only a range of electrode numbers that encompasses those stimulated in this experimental run is shown (electrodes 11 and 7 for tones A and B, respectively). The filled rectangles illustrate when electrical stimulation was present on each electrode.

### 3.3.3 Listeners

Listeners were eight post-lingually deafened adults (6 female, 2 male) who were experienced users of the Nucleus CI24 cochlear implant. All had full electrode insertions. Demographic and clinical data for the participants are shown in Table 3.3, including their open set (sound only) speech recognition score for sentences (recorded within 6 months of the experiment). Two of the listeners (L3 and L7) were tested using only two of the four TRTs (100 and 150 ms).

**Table 3.3:** Demographic and clinical details of the listeners.

| Listener | Age | Gender | Open set speech recognition score (BKB sentences in quiet, % correct) | Pulse rate (pulses per second) in normal daily use | Pulse width (μs) in normal daily use |
|---|---|---|---|---|---|
| L1 | 42 | F | 80 | 1200 | 25 |
| L2 | 48 | F | 76 | 250 | 25 |
| L3 | 61 | F | 91 | 1200 | 25 |
| L4 | 33 | F | 88 | 250 | 25 |
| L5 | 58 | M | 48 | 250 | 25 |
| L6 | 25 | M | 98 | 250 | 25 |
| L7 | 35 | F | 100 | 250 | 25 |
| L8 | 52 | F | 79 | 900 | 25 |

### 3.3.4 Procedure

Listeners were presented with 30-second sequences of alternating stimuli in a repeating ABA-ABA-... format, where A corresponds to a stimulus applied to electrode 11 (in the middle of the electrode array, corresponding to a pure tone at 1938 Hz) and B corresponds to a stimulus on one of the other electrodes, i.e. 1 to 10 and 12 to 21 (N.B.: higher numbers denote more apically placed electrodes). Some larger electrode separations were not included in the test set, because the area of most interest was the smaller electrode separations, located on either side of electrode 11. The electrode used for tone B was selected at random between presentation sequences, but did not vary within each 30-second sequence. In total, five repetitions of each ABA combination were used. This combination of task and sequence structure has previously been found to provide a sensitive and reliable measure of stream segregation in normal-hearing listeners (Cusack and Roberts, 1999; Roberts et al., 2002; Cusack and Roberts, 2004).

### 3.3.5 Task and responses

The purpose of the experiment was explained to the listeners; they were instructed both verbally and in writing (see Appendix 1). Each listener was seated in front of a computer screen and keyboard. They were instructed to press the space-bar on the keyboard whenever their perception of the sequence changed from an 'integrated' (1 stream) to a 'segregated' (2 streams) percept or vice versa. It was explained to the listeners that the 'integrated' percept corresponded with a 'galloping' rhythm, which would disappear for the 'segregated' percept. The instructions given were neutral:

listeners were not instructed to try and hear the stimuli in any particular way. All listeners received training and practice on the task prior to experimental runs. The training used stimuli that were chosen to demonstrate the two 'extremes' of possible percepts, i.e. a) alternating tones presented to electrodes close together, intended to evoke a clearly integrated percept, and b) alternating tones stimulating electrodes that were widely spatially separated, intended to evoke a clearly segregated percept. Care was taken to ensure that listeners understood the task, were able to respond appropriately, and that any task learning that was evident was essentially complete before the experimental runs began. The computer screen displayed their current choice. At the beginning of each 30-second sequence, the screen always read 'Integrated'. This was based on the finding by Bregman (1978) that the default percept is integration, and that the tendency for stream segregation builds up over time. The computer recorded the timing of each key press and from these calculated both the total number of key presses for the sequence and the proportion of time for which the sequence was heard as segregated.

## 3.4 Results

### 3.4.1 Overall results

Figure 3.2 shows the mean segregation reported for each electrode for tone B across all 8 listeners. Overall, a strong effect of electrode number on reported segregation can be seen, and this effect is broadly symmetrical on either side of electrode 11. A within-subjects ANOVA showed a highly significant effect of electrode separation [F(13,65) = 12.145, p<0.001], but there was no significant effect of TRT [F(3,15) = 0.394, p=0.759]. Also, there was no significant interaction between electrode separation and TRT [F(39,195) = 1.040, p=0.416] (N.B: data for some combinations of electrode number and TRT were not tested in all listeners; these cases were treated as missing values in the analysis). As it was considered possible that floor or ceiling effects in some of the data may have affected the analysis, the ANOVA was repeated using arcsine-transformed data, with the intention of reducing the heterogeneity of the data (transformation used: Y'= 2 arcsine( √Y); see Keppel and Wickens (2004), page 155). The results, which confirmed the outcome of the first analysis, were as follows:

Electrode separation: F(13,65)= 12.481, p<0.001; significant

TRT: F(3,15)= 0.563, p=0.648; N.S.

Electrode separation x TRT: F(39,195)=0.841, p=0.734; N.S.

The proportion of time that listeners reported hearing a segregated percept clearly increased as the spatial separation between tone A and tone B increased in terms of electrode number, and the lowest reported segregation was observed when tone B was on electrode 10 or 12 (both immediately adjacent to tone A on electrode 11). Although there were differences in reported segregation for different TRTs at greater electrode separations, it should be noted that there was no systematic effect of varying TRT towards the basal or apical ends of the array.

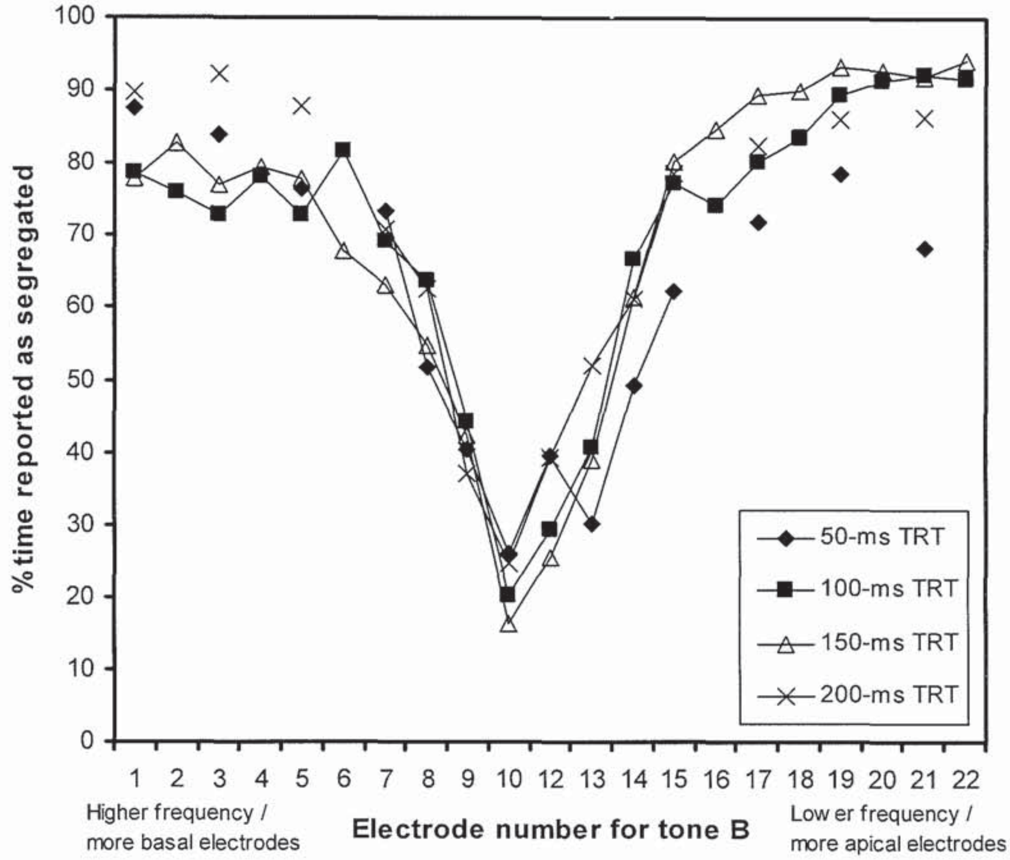


**Figure 3.2.** Results for experiment 1 showing reported segregation in percent (mean data) for all eight listeners. Not all separations between electrode 11 and electrodes near the ends of the electrode array were tested for TRTs of 50 and 200 ms. Data collected in these regions of limited sampling are displayed as isolated points.

### 3.4.2 Individual results

Figure 3.3 shows individual results for each listener and reveals interesting variations. For example, L6 showed a clear tendency to report a segregated percept most of the time, with significant reductions in reported segregation seen only when tone B stimulated electrode 10 or 12, both adjacent to electrode 11. In contrast, at intermediate TRTs (100 and 150 ms), L5 hardly ever reported segregation when tone B stimulated one of the electrodes in the range 7-14, which corresponds to 3-4 electrodes either side of electrode 11.

Inspection of the patterns of key presses recorded for each listener also reveals considerable variability. Figure 3.4 shows the number of key-presses averaged over TRTs of 100 and 150 ms (the TRT values tested in all listeners), because key-press activity did not vary systematically with rate. Most listeners showed a clear tendency towards making only a single key-press, indicating just one change in percept throughout the 30-second listening period. Analysis of 5-second intervals throughout the 30-second period indicated that, when a single key-press occurred, it was almost invariably within the first 5 seconds of the experimental period. Given that the default screen display was 'integrated', this suggests that most listeners tended towards a single judgement of whether they could hear two distinct sounds and, if not, they did not press the space-bar at all. This pattern of responses is not consistent with the perceptual reversals that are usually found for normal-hearing listeners in streaming tasks using long sequences (e.g., Anstis and Saida, 1985; Roberts et al., 2002).

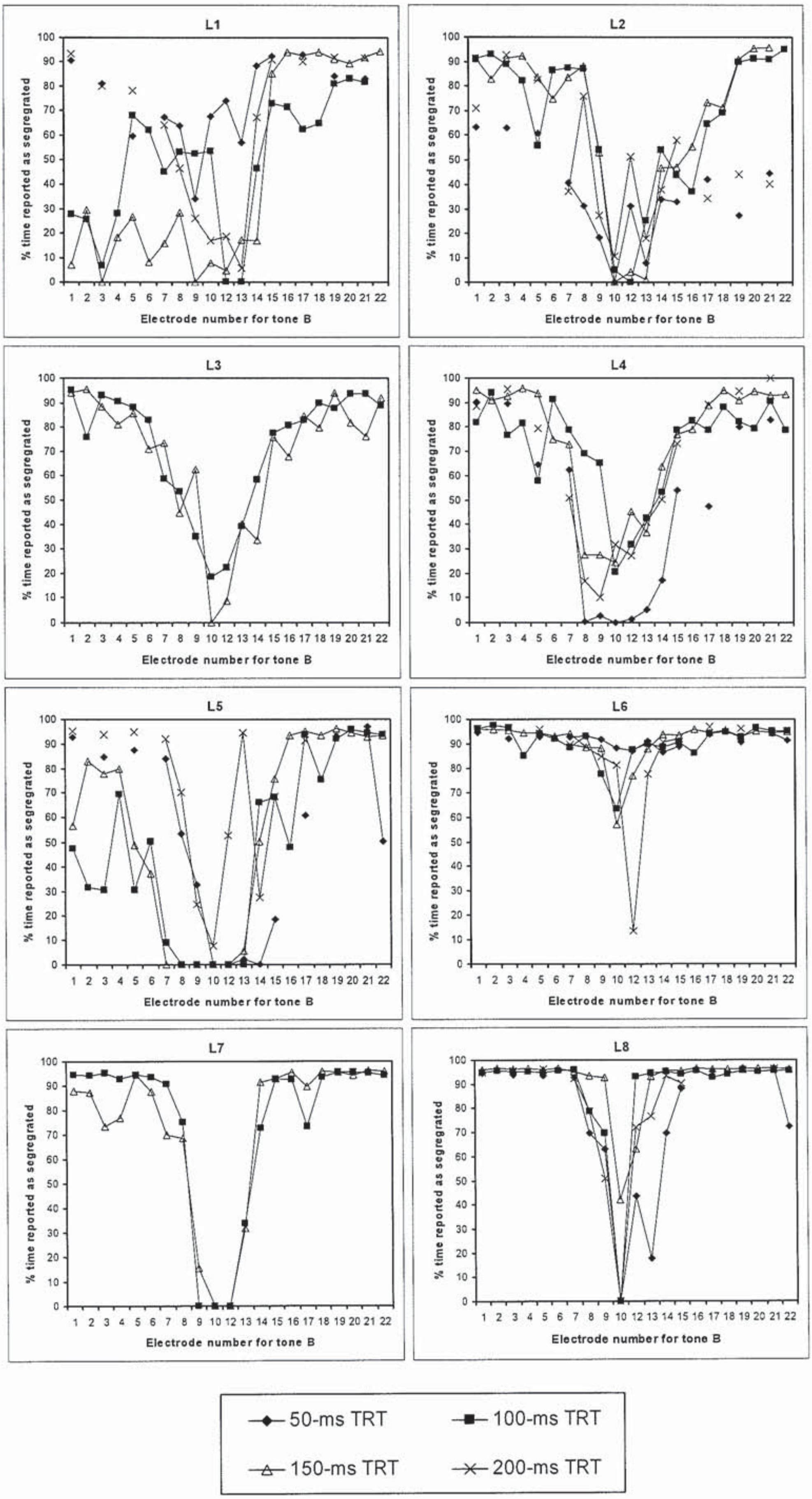


**Figure 3.3**. Reported segregation results for individual listeners.

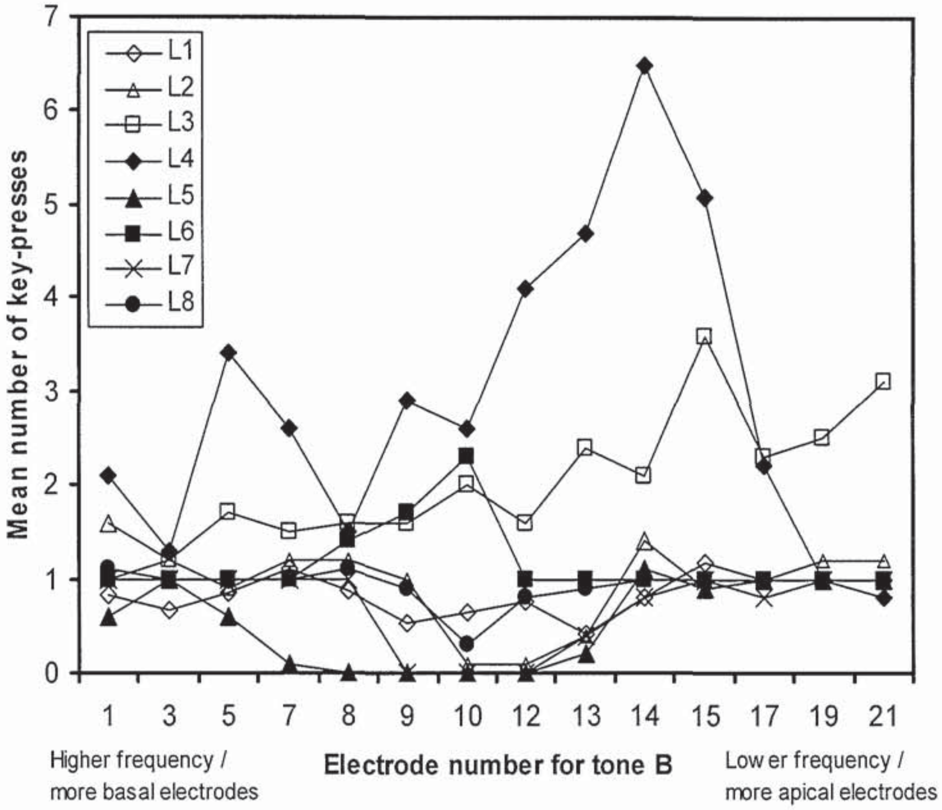


**Figure 3.4.** Key press results for individual listeners.

One listener (L4) did display some relationship between electrode separation and key-pressing activity: she showed a clear tendency towards more key-presses for smaller electrode separations, i.e. for values of tone B corresponding to electrodes between 9 and 16. Indeed, when tone B stimulated electrode 14, the mean number of key-presses for L4 peaked at 6.5. For greater electrode separations (i.e., for values of tone B corresponding to electrode 17 or above, or to electrode 8 or below) her typical mean number of key-presses was low (between 1 and 2), apart from a peak at electrode 5. These results could imply greater ambiguity and perceptual instability for smaller electrode separations, at least for those electrodes more apical to number 11. However, it is notable that L4 did not show any systematic relationship between % segregation and rate of presentation. Indeed, for small electrode separations, % segregation was lowest for the most rapid sequences (50-ms TRT) for this listener. This is the opposite of the typical finding in normal-hearing listeners.

## 3.5 Discussion

The absence of a clear effect of the rate of presentation of tones on the tendency towards reporting a segregated percept, combined with limited evidence of perceptual reversals from the key-press data, indicates that the responses of the cochlear implant listeners in this task were very different from those of normal-hearing listeners. Although it cannot be ruled out entirely, it seems unlikely that electrical stimulation (as opposed to acoustic) would change the basic characteristics of a cognitive process like automatic stream segregation, were it actually taking place in these cochlear implant listeners. Instead, the results for these listeners may have reflected simple channel discrimination judgments rather than auditory stream segregation. Listeners tend to try and follow instructions, and in the absence of any stream segregation, they may instead have interpreted the task as to report when they could hear more than one distinct pitch in the sequence. If this were the case, then little or no effect of the rate of stimulus presentation would be expected over the range used. Neutral instructions per se would not have led to the absence of an effect of TRT (e.g.,see Anstis and Saida, 1985, experiment 1).

The findings, therefore, are broadly consistent with those of Chatterjee et al. (2006), as discussed above. Although there was a clear relationship between *reported* segregation and electrode separation in these listeners, as well as in theirs, the other properties of auditory segregation were not apparent. Chatterjee et al. suggested that the absence of the build-up of streaming that they found in 4 out of 5 of their listeners did not necessarily mean that streaming had *not occurred.* They implied that their listeners may have based their judgements on pitch or quality differences between the stimuli instead of attending to whether streaming was occurring. However, whilst their account cannot be ruled out, it does seem to be a somewhat speculative interpretation of the data.

In order to evaluate our suggested explanation for the pattern of results observed thus far, Experiment 2 investigated pitch ranking across the implant electrode array and compared performance on this task with the results of the stream segregation task. The purpose of experiment 2 was to investigate whether a greater tendency towards reporting a segregated percept in experiment 1 is associated with more reliable (and more confident) pitch judgements.

# Chapter 4: Experiment 2

## 4.1 Introduction

In order to investigate the findings of experiment 1 further, experiment 2 employed a pitch-discrimination task. If the interpretation of the results of experiment 1 described above is correct, namely that the data provide a measure of channel discrimination rather than primitive stream segregation, then overall performance on a more straightforward pitch task, and also individual differences in performance, should correlate with the corresponding segregation scores obtained in experiment 1 (although it should be noted that a positive correlation would also be predicted even if stream segregation had occurred).

In this experiment, listeners judged the pitch (or sharpness/brightness) of each of the other electrodes in the electrode array in comparison with electrode 11. This allowed for direct comparison with the results obtained in experiment 1.

## 4.2 Method

### 4.2.1 Stimuli, task, and procedure

Stimuli were pure tones with identical frequencies to those used in experiment 1 (see Table 3.1). Each tone had a total duration of 120 ms including 10-ms rise and fall times (this corresponds to the tones used in the 200-ms TRT condition in experiment 1). This duration was chosen as it was considered sufficiently long to enable an optimal subjective pitch judgement. As before, all stimuli were equalised for loudness before the experimental trials were run, using the procedure described earlier. Listeners were presented with two consecutive pure tones, with a 1-s silence between them. One of the stimuli, randomly the first or second presented, was always a 1938-Hz pure tone that stimulated channel 11 (tone A). The other stimulus (tone B) was selected at random to stimulate any one of the other electrodes (1 to 10 or 12 to 22). The listeners were instructed to listen to both sounds, and then indicate which of the sounds (first or second) sounded higher pitched, or brighter/sharper in timbre, by pressing either key 1 or key 2 (see Appendix 2 for written instructions). No feedback was provided after each response. After responding in this way, listeners were asked to indicate how confident they were in their choice, on a scale from 1 to 5, by pressing the appropriate key. The values of these keys corresponded to these descriptions: 1 = Very unsure, really guessing; 2 = A little bit confident; 3 = Moderately confident; 4 = Very confident, but not certain; 5 = Absolutely confident. In total, ten repetitions of each electrode pairing, in both possible orders, were presented (in random order) to each listener. Both pitch ranking and confidence ratings were recorded.

### 4.2.2 Listeners

Listeners were 6 post-lingually deafened adults who were experienced users of the Nucleus CI24 cochlear implant. All 6 had participated in experiment 1. The other two listeners from experiment 1 (L3 and L8) were unavailable.

### 4.3 Results

Figure 4.1 shows the pitch-ranking results for all 6 listeners. Considerable variability is apparent across listeners in the slope of the function describing electrode number vs. pitch ranking in relation to electrode 11. As expected, all the curves converged on a probability of 50% when both stimuli were on electrode 11. Most listeners show the expected monotonic relationship between electrode number and perceived pitch: higher electrode numbers (more apical to electrode 11) are more likely to elicit a lower pitch (and so the probability of being reported as higher in pitch than electrode 11 falls). For the lower electrode numbers (more basal than electrode 11) the converse is true. In a perfect situation, every electrode higher in number than 11 would always be reported as lower pitched than 11, and all those lower than 11 would always be reported as higher pitched than 11. For some listeners, (e.g., L6) this is essentially the case. However, the situation is less clear-cut for many other listeners, for whom the responses suggest much more ambiguity in their pitch perceptions. In particular, L2 displayed non-monotonic responses in her pitch judgements, with no clear relationship to channel number. L4 showed the expected trend when responding to more basal electrodes, i.e. generally reporting them as higher in pitch than electrode 11, while for more apical electrodes her responses were far more ambiguous.

Similar variability can be seen in the confidence rating scores shown in Figure 4.2. Some listeners (e.g., L6) reported a high degree of confidence in the vast majority of their pitch judgements, which only fell to near 1 (corresponding to guessing) when the spatial separation from electrode 11 was small or absent.

Others (e.g., L2) required greater electrode separations to be confident in their pitch judgements. The overall mean confidence score (collapsed across all electrode separations) ranged across listeners from 3.6 to 4.6. Generally, those listeners who reported the highest confidence showed the sharpest tuning in their pitch ranking judgements. However, L2, who showed apparently rather disordered pitch perception (as discussed above), reported quite high confidence in her judgements except for small electrode separations. The variance in confidence scores may partly reflect individual differences in overall confidence in responses to psychophysical tasks.

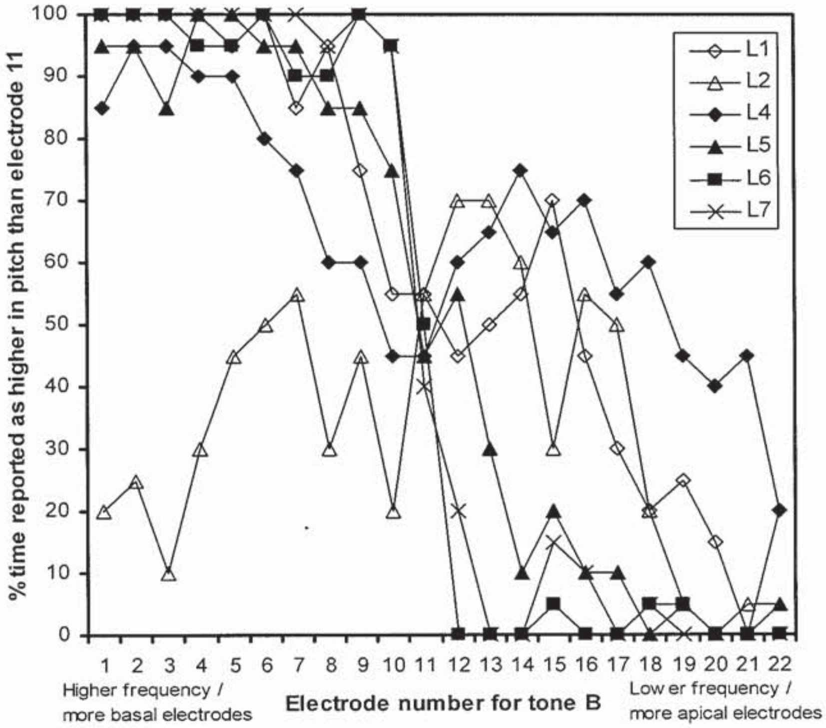


**Figure 4.1** Pitch ranking results for individual listeners.

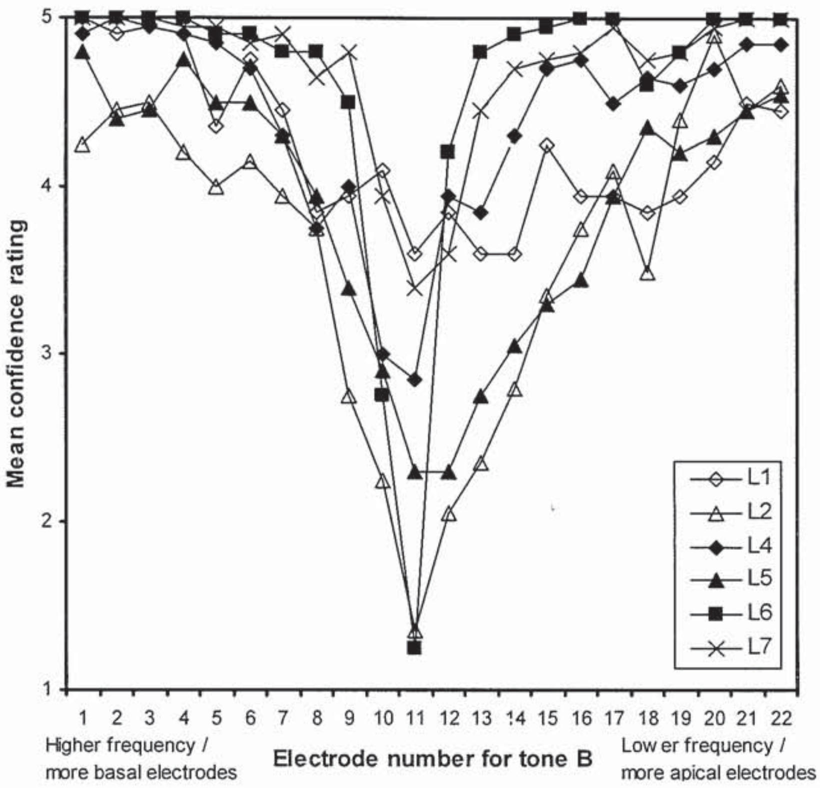


**Figure 4.2.** Reported confidence ratings for individual listeners.

## 4.4 Comparison of segregation and pitch ranking judgments

The results of experiment 1 suggested that these implant listeners experienced automatic stream segregation only rarely, if at all, even when tones A and B stimulated widely separated electrode pairs. Instead, it could be that implant listeners may often have defaulted to judging how many pitches (or timbral brightnesses) they heard in a test sequence. If this is so, then a relationship might be anticipated between a tendency towards reporting a segregated percept in experiment 1 and discriminating pitch reliably (and with confidence) in experiment 2.

The data for the six individual listeners who took part in both experiments were transformed to facilitate qualitative and quantitative comparisons of the different judgements. For each channel separation, the transformations used were as follows: (i) reported segregation scores were averaged for TRTs of 100 and 150 ms (the values tested in all listeners); (ii) the percentage of responses for which each electrode was reported as higher in pitch than electrode 11 was converted into an absolute difference from chance (50%) and re-scaled from 0 to 100 (i.e., absolute difference x 2); (iii) the confidence ratings were re-scaled from the 1 to 5 range that listeners used into a 0 to 100 scale (i.e., 25 x (rating – 1)). An absolute difference from chance was used as an approximate measure of discriminability when transforming the pitch rankings, because it was assumed that the key factor is electrode discriminability rather than the perceived direction of the pitch difference. Furthermore, a monotonic relationship between electrode number and perceived pitch height cannot be assumed for all implant listeners.

Figure 4.3 shows the relationship between reported segregation, pitch discrimination, and the confidence of listeners in their pitch rankings. In general, with the exception of L1 in the basal direction and L4 in the apical direction, there is a similarity in the pattern of individual listeners' results across these three measures. For example, L6 showed a strong tendency towards reporting segregation except for those electrodes closest to electrode 11. He also showed nearly perfect pitch ranking performance in experiment 2 and reported the highest overall confidence (mean = 4.6). In comparison, L5 showed broader tuning on all three measures.

The overall relationship between reported segregation and confidence in pitch ranking is closer than that between reported segregation and pitch discrimination. For example, L4 reported high segregation and high confidence in her pitch judgements when tone B stimulated electrodes towards the apical end of the array, despite her poor discrimination of pitch for these electrodes relative to electrode 11.

Correlations between the reported segregation scores and a) the transformed pitch rankings and b) the transformed confidence ratings for each listener and overall (i.e. pooled data from all listeners) were calculated (Table 4.1). A separate correlation was calculated for each of the two directions of electrode separation, i.e. basal (electrodes 1 to 10) and apical (electrodes 12 to 22). Five out of the 12 correlations calculated between reported segregation and pitch ranking for individual listeners were significant and positive. The overall correlation was significant and positive in the apical direction, accounting for about 30% of the variance, but it was not significant in the basal direction. Nearly all (11/12) of the correlations between reported segregation and the confidence ratings for individual listeners were significant and positive, and the overall correlations were significant and positive in both apical and basal directions, accounting for about 60% and 20% of the variance, respectively.

These findings confirm the observation that there is a closer relationship between reported segregation and reported confidence in pitch ranking judgements than with pitch discrimination per se. It should be noted that, for L6, the correlation between his reported segregation scores and transformed pitch rankings was low and not significant, despite his similar and high scores on these two measures. This was presumably the result of a ceiling effect, because many of the values for both measures were close to 100%. In contrast, L6 showed a very high and significant correlation between his reported segregation scores and confidence ratings.

**Table 4.1.** Correlations (adjusted $r^2$ values) between reported segregation scores and (a) transformed pitch scores and (b) transformed confidence scores. Separate correlations were calculated for each direction of electrode separation (i.e.: basal, electrodes 1 to 10, and apical, electrodes 12 to 22). Correlations that are statistically significant ($p<0.05$) are shown in bold text.

| Listener | Direction of electrode separation | Reported segregation and transformed pitch scores | | Reported segregation and transformed confidence scores | |
|---|---|---|---|---|---|
| | | $r^2$ | p values | $r^2$ | p values |
| L1 | Basal | 0.025 | 0.664 | 0.341 | 0.076 |
| | Apical | **0.531** | **0.011** | **0.495** | **0.016** |
| L2 | Basal | 0.001 | 0.943 | **0.834** | **<0.001** |
| | Apical | **0.398** | **0.038** | **0.895** | **<0.001** |
| L4 | Basal | **0.851** | **<0.001** | **0.937** | **<0.001** |
| | Apical | 0.034 | 0.588 | **0.889** | **<0.001** |
| L5 | Basal | **0.411** | **0.046** | **0.628** | **0.006** |
| | Apical | **0.835** | **<0.001** | **0.830** | **<0.001** |
| L6 | Basal | 0.034 | 0.612 | **0.955** | **<0.001** |
| | Apical | 0.041 | 0.550 | **0.563** | **0.008** |
| L7 | Basal | 0.254 | 0.138 | **0.596** | **0.009** |
| | Apical | 0.229 | 0.137 | **0.853** | **<0.001** |
| Overall | Basal | 0.008 | 0.498 | **0.229** | **<0.001** |
| | Apical | **0.302** | **<0.001** | **0.619** | **<0.001** |

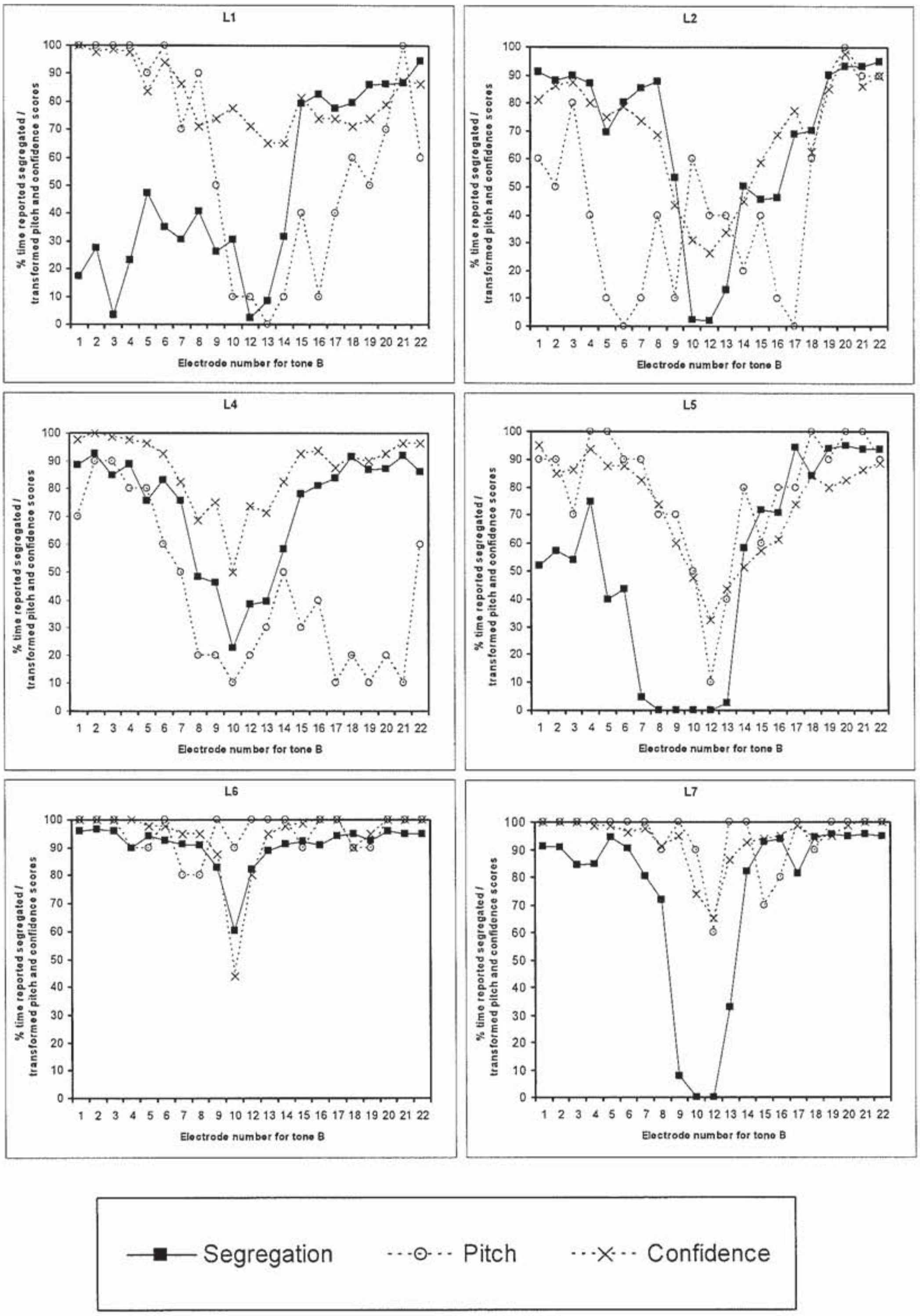


**Figure 4.3**. Comparison of individual results for reported segregation, transformed pitch rankings, and transformed confidence ratings.

## 4.5 Discussion

The overall relationship between the measures obtained in experiments 1 and 2 is broadly consistent with the proposal that the results of experiment 1 primarily reflected the channel discrimination abilities of these listeners. However, correlation does not prove causation, and there is an alternative explanation of this relationship that merits consideration. Specifically, the correspondence between reported segregation and pitch ranking may have arisen simply because implant listeners who are more capable of perceptually mapping pitch are also better able to segregate different frequencies into separate perceptual streams. Although this account cannot be ruled out, two observations are more compatible with the interpretation outlined. First, as already discussed, these implant listeners have shown an absence of the rate and ambiguity effects characteristic of streaming studies in normal-hearing listeners. Second, three of the six listeners who took part in both experiments (L2, L4, and L6) showed an exceptionally close parallel between reported segregation and their transformed confidence ratings for pitch ranking, with large and highly significant correlations between the two measures (see Figure. 4.3 and Table 4.1). Indeed, the data sets almost overlap for two of these listeners. This indicates that these listeners were reporting segregation close to the point at which they began to become confident in their pitch rankings.

In summary, the results of experiments 1 and 2 demonstrate that clear evidence of automatic stream segregation is difficult to produce with this kind of task in implant listeners; reported segregation seems to relate mainly to listeners' confidence in their judgements about pitch differences between stimuli on different electrodes. A more objective approach, using a task in which segregation impairs accuracy of performance despite the efforts of listeners, may be more effective and is explored in experiment 3.

# Chapter 5 : Experiment 3

## 5.1 Introduction

### 5.1.1 Judgements of order and relative timing as streaming measures

In experiment 1, the task required listeners to report their experience of whether they heard one or two streams. It therefore provided a direct, although subjective, measure of *reported* segregation. However, although a significant relationship between reported segregation and electrode separation was found, there was no significant effect of presentation rate, and little evidence of the perceptual instability for intermediate electrode separations that would have been predicted if stream segregation was indeed taking place. Thus, the results did not provide convincing evidence of auditory stream segregation in cochlear implant listeners. An alternative approach is to use a task in which perceived segregation is not reported directly, but where stream segregation, if it occurs, will be expected to adversely affect performance. As discussed in the introduction, the perceptual properties of sounds are computed within streams and not across them. For example, as discussed earlier, performance is impaired in tasks that require the listener to count or report the order of sequences of sounds if the sounds split into separate auditory streams (Warren et al., 1969; Bregman and Campbell, 1971).

Other examples of tasks where performance would be expected to be impaired by stream segregation are those that require judgements about the relative timing of sounds. This was investigated in one of the experiments described by van Noorden (1975). He used continuous sequences of alternating AB tones, where the TRT for tone B was either 1% smaller or 1% larger than for tone A; as a result, the B tones gradually shifted from the midpoint between the A tones. Listeners reported when they perceived that the B tones were no longer at the temporal midpoint; when this occurred, the relative position of the A and B tones was noted, and the TRT for tone B was changed from +1% to -1% relative to tone A or vice versa. Using this tracking method, he measured the just perceptible displacement ($\Delta T$) of the B tones from the midpoint between the neighbouring A tones in the sequence. There was a very clear dependence of $\Delta T$ on the tone interval in semitones between tones A and B ($\Delta T$ increased significantly with increased frequency separation). He then measured the values of $\Delta T$ for short tone sequences (i.e. two tone pairs, AB…AB). There was again some dependence of $\Delta T$ on the A/B frequency separation, but to a much lesser degree than with the continuous sequences. Interestingly, overall performance was better with the short than with the continuous tone sequences; the smallest value for $\Delta T$ was when tones A and B were at the same frequency, and was 5 ms for the short sequence while higher for the long sequences. Finally, he measured $\Delta T$ for a 3-tone (ABA) sequence by omitting every other ABA group in the continuous sequences, so that the remaining ABA triplets were perceived as separate, short tone sequences. In this case, there was very little dependence of $\Delta T$ on A/B frequency separation (e.g., for an A-B interval of 6 semitones, the $\Delta T$ for short, 3-tone sequences was around 20 ms compared with around 40 ms for a continuous sequence). This was evidence that temporal discrimination is influenced by the build-up of stream segregation over time. He concluded that all these results supported the view that temporal position *can only be perceived properly* when there is temporal coherence (integration).

The idea that stream segregation produces a loss of information regarding the sequential relations of two alternating sounds was also explored by Dannenbring and Bregman (1976). They asked listeners to judge the extent to which two different sounds alternating rapidly overlapped in time. They showed that the degree of reported overlap more than doubled when listening to tones expected to perceptually segregate (1000 Hz and 2600 Hz) compared with a 'non segregating' pair (1000 Hz and 1100 Hz). They took this as evidence for a process of perceptual decomposition, which dissociates the temporal features of events in different perceptual streams; Bregman (1990) pointed out that the relative timing of auditory events is a relevant perceptual property only if those events are related.

A similar approach to van Noorden's was used by Vliegen et al. (1999) in an experiment that investigated the role of spectral and periodicity cues in auditory stream segregation (see chapter 1 for more details).

Cusack and Roberts (2000) used a temporal discrimination task to follow up the results of an interleaved melody task that investigated the effects of timbral contrasts on stream segregation. They argued that it was not clear whether the results of their first task were due to primitive (automatic) stream segregation or to schema-based selection; the use of a schema to select sounds with a different timbre could have led to the observed differences in performance. In their second experiment, good performance depended on the integration of sounds in a sequence; selection of sounds by 'top-down' schema-driven processes, under attentional control, cannot improve performance on this kind of task, and so changes in threshold should be determined primarily by involuntary (primitive) stream segregation.

Their stimuli consisted of sequences of 28 alternating high and low frequency sounds. For the first 8 sounds, the higher sounds were always halfway in time between the two neighbouring low frequency sounds (isochronous). In half of the trials, the isochronous rhythm was continued throughout the whole sequence. In the other half, after the first eight sounds, the onset of the higher sounds was progressively delayed over the next 12 sounds, and the resulting rhythm was maintained for the final 8 sounds. The initial *isochronous* portion of the target stimulus was long enough to allow significant build-up before the anisochrony was introduced. In this case, the threshold delay was not measured; the size of the temporal offset used was selected on the basis of pilot work so that the change in rhythm was only salient if the sequence was heard as a single stream. Thus, the ability to detect the anisochronous rhythm was used to indicate an absence of segregation. This allowed the measurement of threshold frequency separations between the high and low sounds for four different combinations of pure tones and narrow-band noise. The results showed that differences in timbre can enhance stream segregation even when the listener is trying to integrate the sounds, implying that a primitive grouping process was the cause of the observed results.

Temporal discrimination was used again as a measure of stream segregation by Roberts et al. (2002). Their test sequences were similar to those used by Cusack and Roberts (2000), except that delay was controlled rather than A-B frequency separation. Using this method, they measured the threshold delay for various complex tones, all comprising unresolved harmonics of F0 = 100 Hz, and showed that dissimilarity in either passband or phase (the latter creating differences in pitch and/or timbre) increased discrimination thresholds. They then compared the results of the temporal discrimination task with a task that required subjective judgements of stream segregation, in which listeners judged moment-by-moment grouping of long ABA..ABA sequences. They showed that the factors that increased thresholds in the temporal task were very similar to those that increased perceived segregation. Thus, each task was in effect a different method of measuring the same underlying process. Temporal discrimination tasks continue to be used as a measure of stream segregation (e.g. Roberts et al., 2008, in press).

In summary, temporal discrimination has been used to provide an indirect measure of primitive stream segregation in a number of studies with normally hearing listeners; the studies by Cusack and Roberts (2000) and Roberts et al. (2002) showed that results obtained in this way can correspond well with tasks requiring either selection of target sounds from distractors or subjective reporting of segregation.

One important caveat regarding this approach is worthy of note at this point. Even for two isolated pure tones (i.e.not part of a sequence), for which one would expect little build-up in the tendency for stream segregation, thresholds for detecting a temporal gap between them rise for normal-hearing listeners as the frequency separation is increased. For example, Grose et al. (2001) measured the gap detection thresholds for pure tones at 4000 Hz and 2285 Hz, and 10 Hz and 20 Hz FM tones and 10 Hz and 20 Hz AM tones (carrier frequency = 4000 Hz), in various combinations pre- and post- gap. Gap detection thresholds were significantly higher when the frequencies of the pure tones either side of the gap differed (4000 Hz vs 2285 Hz) than when they were the same. However, other perceptual dissimilarities (i.e. presence/absence of FM or different rates of FM) did not lead to worsened gap thresholds. Similarly, Neff et al. (1982) reported that, for normally hearing listeners listening to short sequences of tones of various durations and frequencies, gap discrimination became poorer as the difference between the frequencies of the tones bounding the gap increased, but was not influenced by changes in sequence length and rate. Clearly, the effects of frequency differences between tones on performance in gap discrimination tasks do not always reflect increased stream segregation.

Gap detection thresholds have also been shown to rise as the physical separation of stimulated electrode pairs is increased in cochlear implant listeners; Hanekom and Shannon (1998) used this as a measure of electrode interaction, *not* of stream segregation. Therefore, an increase in temporal discrimination thresholds with increased frequency separation between two stimuli can not *in itself* be taken as firm evidence of stream segregation, as it may occur anyway. Therefore, in order to demonstrate convincingly that changes in threshold reflect changes in stream segregation, it is necessary to observe a *greater dependence* of the temporal judgements on frequency separation in situations where stream segregation is expected to be greater, i.e. when there has been sufficient time for segregation to build up, than when it has not. This was demonstrated by van Noorden (1975) in his comparison between long sequences and short ABA triplets.

### 5.1.2 Temporal discrimination as a streaming measure in implant listeners

An example of the application of this approach in cochlear implant listeners is the experiment reported by Hong and Turner (2006). They used identical sequences of tones to those described by Roberts et al. (2002), with 7 normally hearing (NH) and 8 cochlear implant (CI) listeners. Stimuli were presented in the sound field via loudspeaker - at 90 or 95 dB SPL for the implant listeners, who used their normal speech processors and 'Maps' (speech processing strategies and stimulus levels), and were allowed to adjust their volume or sensitivity settings if required. In the tone sequences, tone A was set at either 200, 800 or 2000 Hz, while the frequency of tone B was set at a ratio ((B-A)/A) of either 0 (i.e. A=B), or 0.01, 0.1, 0.5, 1.0 or 3.0. Using an adaptive 2I-2AFC task, they measured the threshold delay for tone B for this range of A/B frequency separations and for each 'base frequency' for tone A. The results were normalised relative to the threshold delay for a Weber fraction = 0, i.e. A=B. They found considerable variability in the slopes of the threshold delay vs. frequency separation functions for the CI listeners. They then converted the A-B frequency differences into electrode separations, based on the frequency allocations to each electrode in the CI listeners' 'Maps', and reported significant linear regressions between (normalised) threshold delays and electrode separations for some (though not all) of the listeners.

In an attempt to determine whether the observed relationship between frequency or electrode separation and threshold delay truly reflected stream segregation, they repeated the task with 3 NH and 3 CI listeners using isolated ABA tone triplets. Their hypothesis was that if their listeners performed worse on the 'streaming task', i.e. with full-length tone sequences, than on the 'short rhythm', i.e. tone triplet task, this could be taken as evidence that the long-sequence task measured stream segregation. They found that for the 3 CI and 3 NH listeners, baseline performance (tone A = tone B) was worse (higher thresholds delays) for the ABA triplets than the full-length tone sequences, while the reverse was the case when the ratio between the frequencies of tone A and tone B was 1:1.5 (i.e. A and B = 200/300 Hz, 800/1200Hz and 2000/3000 Hz, respectively). Therefore, they in effect found an interaction between tone sequence length and frequency separation, which they asserted was evidence that the task used was a genuine measure of stream segregation. They went on to report correlations between their normalised thresholds in the streaming task and speech reception thresholds in noise.

While potentially demonstrating auditory stream segregation and its relationship to speech recognition in noise in CI listeners, the results reported by Hong and Turner (2006) should be treated with some caution, and there are a number of caveats relating to their experimental methods. Firstly, stimuli were presented in the sound field via loudspeaker, and so the effects of each listener's speech processing strategy on the resulting patterns of electrical stimulation are not clear. Indeed, a mixture of implant systems and speech-coding strategies was evident amongst the participants in each of their experiments, and they were allowed to adjust the volume or sensitivity levels of their speech processors. Therefore, the precise characteristics of the electrical stimuli that were delivered are unknown. Secondly, frequency separation was the experimental parameter that was manipulated and this was only converted into electrode separation afterwards. This inevitably leads again to some uncertainty about the actual stimuli that were delivered to the listeners' electrodes. Finally, only 3 CI listeners (one user of the Nucleus implant and two of the Clarion implant), picked at random from the group used in the first experiment, participated in their second experiment, which is important for any conclusion that the results with the full-length sequences actually reflected stream segregation. Results from a larger number, based on more precisely controlled stimuli, would be required to provide convincing evidence of stream segregation in the majority of implant listeners.

## 5.2 Overview of Experiment 3

In order to investigate stream segregation in CI listeners using a temporal discrimination task, experiment 3 employed the same task and stimulus sequence structures as did Roberts et al. (2002), but also used ABA tone triplets. In contrast to Hong and Turner (2006), the stimuli used were presented via direct input to the listeners' speech processors, as in experiments 1 and 2. Again, the frequencies of the tones used were such that they fell in the centre of the passband for the intended electrodes, and the precise pattern of electrical stimulation was verified using the same method as described for experiments 1 and 2. The speech processor was programmed so that it could deliver stimulation on only one channel at a time. An experimental speech processor was used, and listeners had no control over the volume or sensitivity level. All stimuli were presented at C-level (Maximum Comfortable Level) and were loudness balanced (see chapter 3 for details of this procedure). Thus, the relationship between electrode separation and threshold delay was measured directly, both for the full-length tone sequences and for short ABA triplets, in a larger group of CI listeners all using the same type of Nucleus implant. An interaction between tone sequence length and electrode separation, across a range of AB frequency separations, would provide more convincing evidence that stream segregation had indeed mediated the results; in this case, one would expect to see a steeper electrode separation/threshold delay function for the full tone sequences than for the triplets.

## 5.3 Method

### 5.3.1 Stimuli and tasks

In task 1, the full-length sequences of alternating tones had an identical configuration to those used by Roberts et al. (2002) (see figure 5.1). Each sequence lasted 2.4 seconds and consisted of 12 A..B cycles of alternating pure tones (24 tones in total). Tone A was always at 1938 Hz and stimulated electrode 11 (e11). Tone B was at a frequency set to stimulate one of either 6 electrodes more basal to e11 (e1, e3, e5, e7, e9 or e10) or one of 7 electrodes more apical to e11 (e12, e13, e14, e15, e17, e19 or e21), or e11 itself. As the task was time-consuming, even-numbered electrodes spatially distant from e11 were omitted, while for the section of 7 electrodes in the area of interest centred on e11, every electrode was included (see table 5.1). The electrode used for tone B was selected from the test set at random between presentation sequences, but did not vary within each sequence. The duration of each tone in both tasks was 60 ms, including 10-ms onset and offset ramps. In the standard (regular) sequences, each B tone was presented at the exact midpoint in time between the A tones preceding and following it, with a 40-ms silent interval between each tone, giving an onset-to-onset time between consecutive tones of 100 ms. The test (irregular) sequences were configured in the same way as the reference for the first 6 AB cycles, but tone B was progressively delayed in equal steps over the next 4 cycles, and the cumulated delay was maintained for the final 2 cycles. Delays of 0, 10, 20, 30 or 40 ms were used and randomised across presentations; 40 ms was the maximum delay possible without temporal overlap between tones A and B. The method of constant stimuli was used instead of an adaptive procedure, as it was considered important that the CI listeners regularly experienced trials in which they were able to perform well. Pilot work had shown that when the task was consistently difficult, with frequent incorrect responses, listeners found it hard to maintain concentration and remain "on task", which could have led to impaired performance.

**Figure 5.1**. Schematic representation of stimulus sequences used in task 1. Each tone is depicted by a short solid line; tones heard as belonging to the same auditory stream are linked by dotted lines. The upper and lower panels represent a sequence heard as integrated (one stream) or segregated (two streams) respectively. (Adapted with permission from Roberts et al, 2002).

For task 2 (triplets), short A..B..A sequences consisting of 3 tones only were used. Characteristics of the A and B tones were identical to those used for the full-length sequences described above and shown in table 5.1. In the standard (regular) triplet, the B tone was set to occur at the exact midpoint in time between the A tones preceding and following it, with a 40-ms silent interval between each tone. In the test (irregular) triplets, an onset delay of 0, 10, 20, 30 or 40 ms was applied to the central B tone; the delay value was randomised across presentations. Thus, the ABA triplets with delays were identical to the last 3 tones in the full-length test sequences. Exact timing characteristics of the test sequences used are shown in table 5.2.

**Table 5.1**. Frequency characteristics of the tones used and their relation to the implant channels.(N.B: values correspond to those shown in Table 3.2)

| **Channel/electrode number** | **Lower frequency boundary (Hz)** | **Upper frequency boundary (Hz)** | **Channel centre frequency and frequency of the pure tone stimuli (Hz)** |
|---|---|---|---|
| 1 | 6938 | 7938 | 7438 |
| 3 | 5313 | 6063 | 5688 |
| 5 | 4063 | 4688 | 4375 |
| 7 | 3063 | 3563 | 3313 |
| 9 | 2313 | 2688 | 2500 |
| 10 | 2063 | 2313 | 2188 |
| 11 | 1813 | 2063 | 1938 |
| 12 | 1563 | 1813 | 1688 |
| 13 | 1313 | 1563 | 1438 |
| 14 | 1188 | 1313 | 1250 |
| 15 | 1063 | 1188 | 1125 |
| 17 | 813 | 938 | 875 |
| 19 | 563 | 688 | 625 |
| 21 | 313 | 438 | 375 |

**Table 5.2**. Characteristics of stimuli for task 2 (triplets).

| Onset delay applied to tone B (ms) | Silent interval between 1st tone A and tone B (ms) | Silent interval between tone B and 2nd tone A (ms) |
|---|---|---|
| 0 | 40 | 40 |
| 10 | 50 | 30 |
| 20 | 60 | 20 |
| 30 | 70 | 10 |
| 40 | 80 | 0 |

The set up and method of stimulus control was identical to that used in experiments 1 and 2 (see chapter 3 for full description). Stimuli were verified by generation of electrodograms, as described in chapter 3.

### 5.3.2 Listeners

Listeners were 8 adult, post-lingually deafened, experienced cochlear implant users of the Nucleus CI24 implant system. Four of them had previously taken part in experiments 1 and 2. Table 5.3 shows their demographic and other details. Six listeners participated in both the full-length tone sequences task and the 'triplets' task. The remaining 2 participated only in the first task and were unavailable to take part in the 'triplets' task.

**Table 5.3**. Demographic and other details for implant listeners

| Listener | Age | Gender | Open set speech recognition score (BKB sentences in quiet, % correct) | Pulse rate (pulses per second) in normal daily use | Pulse width (µs) in normal daily use |
|---|---|---|---|---|---|
| L2 | 48 | F | 76 | 250 | 25 |
| L3 | 61 | F | 91 | 1200 | 25 |
| L5 | 58 | M | 48 | 250 | 25 |
| L6 | 25 | M | 98 | 250 | 25 |
| L9 | 46 | M | 99 | 1800 | 25 |
| L10 | 61 | F | 72 | 900 | 25 |
| L11 | 45 | M | 92 | 900 | 25 |
| L12 | 69 | F | 93 | 1200 | 25 |

### 5.3.3 Procedure

A 2I-2AFC procedure was used in both tasks. In both task 1 (full length sequences) and task 2 (triplets), all possible combinations of electrode pairings and onset delay for tone B were used in random order. Thus, 4 possible onset delays (10, 20, 30, or 40 ms) x 14 electrode pairings = 56 combinations. Stimulus presentation was initiated by the listener using a key press, and there was a 1-s silence between each interval. In each trial block, the test sequence occurred at random once in the first interval and once in the second, for each delay/electrode-pair combination, giving a total of 112 presentations per block. A total of 10 trial blocks was completed, giving 20 presentations of each onset delay/electrode-pair combination in random order.

The purpose of the experiment was explained to the listeners; they were instructed both verbally and in writing (see Appendix 3). Each listener was seated in front of a computer screen and keyboard. They were instructed to listen to both sequences of tones (or both tone triplets), then respond by pressing key 1 or 2 on the keyboard to indicate which sequence, first or second, was irregular in rhythm. At the start of each trial, the computer screen displayed the instruction:

**PRESS 1 IF FIRST SEQUENCE IRREGULAR**
**PRESS 2 IF SECOND SEQUENCE IRREGULAR**

Feedback was provided on the computer screen based on the accuracy of the responses (i.e. 'CORRECT' or 'INCORRECT'). All listeners received training and practice on both tasks prior to experimental runs. For this, the stimulus expected to be most easily discriminable was used, i.e. tone A=tone B and with the maximum initial delay on tone B of 40 ms (the threshold delay when tones A and B were at the same frequency was less than 40 ms for all the CI listeners for both the full-length sequences and the triplets, so this condition ensured that the task was achievable). Training was continued where possible to a point where each listener understood the task and what to listen for, responded confidently and appropriately, and was achieving very good or perfect performance. Care was taken to ensure that any task learning was essentially complete before the experimental runs began.

### 5.4 Results

The computer software calculated the total number of correct responses for each combination of delay and electrode separation. This was then converted into percentage correct scores. The threshold onset delay of tone B was calculated for each electrode from the percentage correct scores for each delay value tested using a Logistic function to fit the data and estimate the delay equivalent to 75% correct. This provided a good fit to the data in over 90% of cases. In the cases of floor or ceiling performance, thresholds of 40 ms or 10 ms were allocated, respectively.

#### 5.4.1 Mean data

The geometric mean data for the 6 implant listeners who completed both tasks are shown in figure 5.2. This shows clearly the overall trend towards higher thresholds for the triplets than for the full-length sequences for all electrode positions, and the obvious dependence on electrode separation from e11 (higher thresholds associated with greater electrode separation from e11) for both sequence durations.

A within-subjects ANOVA performed using log-transformations of the threshold data revealed a highly significant main effect of electrode number for tone B: $F(13,65)=15.056$, $p<0.001$. There was also a significant effect of sequence length (full length vs. triplets): $F(1,5)= 7.652$, $p=0.040$, but there was no significant interaction between electrode number and sequence length: F(13,65)=1.104, p=0.372. In order to reduce the possibility that ceiling effects had distorted the outcomes of this analysis, the ANOVA was repeated excluding the threshold data for electrodes 1, 3, 19 and 21 (at each extreme end of the electrode array). This analysis confirmed a significant effect of electrode number: F(9,45)=11.333; p<0.001 and of sequence length: F(1,5)= 9.025; p=0.030. Again, there was no significant interaction between sequence length and electrode number: F(9,45)= 1.109; p=0.376.

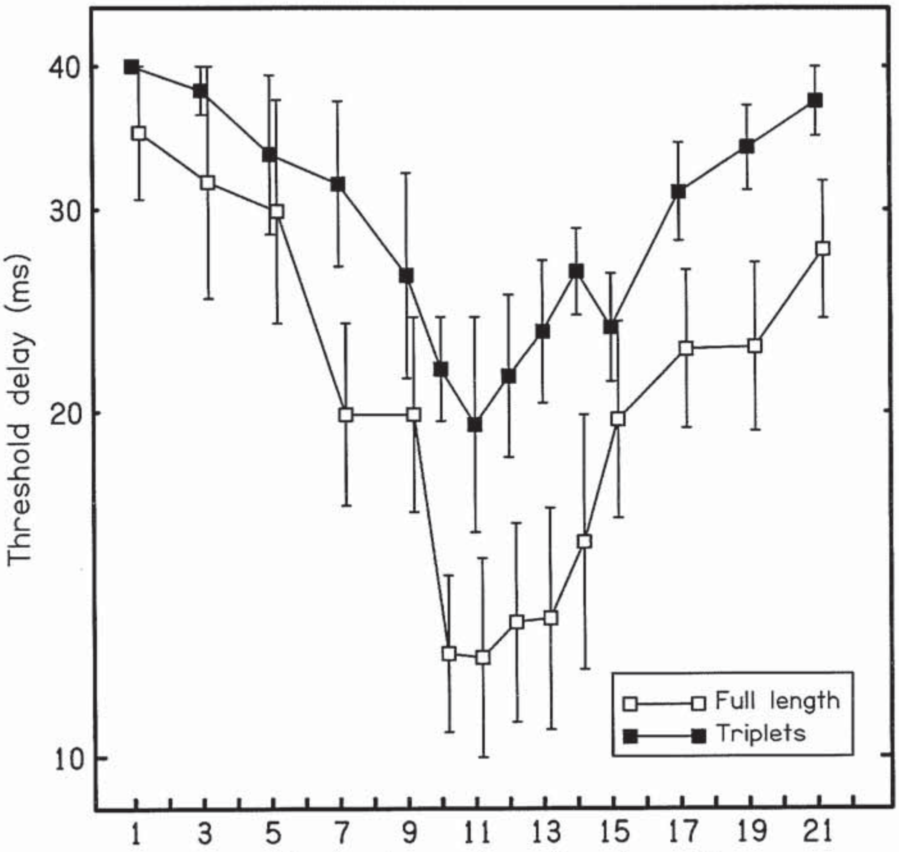


**Figure 5.2**. Geometric mean threshold delays for 6 implant listeners who participated in both task 1 (full-length sequence) and task 2 (triplets). For the purposes of calculation of means, thresholds above the maximum of 40 ms or below the minimum of 10 ms have been allocated values of 40 and 10 respectively. Standard errors for each mean are shown by vertical lines. N.B:

means and error bars for the two sequence lengths have been slightly displaced for clarity.

In order to explore further the relationship between threshold delay and electrode separation (between tones A and B) separately for the basal and apical direction (relative to e11), the ANOVA was repeated for each half of the electrode array; the results were as follows:

i) **Basal direction** (e1 to e11):

Electrode separation: F(6,30)=22.782, p<0.001; *Significant*

Sequence length: F(1,5)= 3.682, p=0.113; N.S.

Sequence length x electrode separation: F(6,30)=1.306, p=0.285; N.S.

*Repeat analysis, excluding extreme values for tone B (e1 and e3):*

Electrode separation: F(4,20)=24.201=p<0.001; *Significant*

Sequence length: F(1,5)= 4.652, p=0.084; N.S.

Sequence length x electrode separation: F(4,20)= 1.166; p=0.355; N.S.

ii) **Apical direction** (e11 to e21):

Electrode separation: F(6,30)=20.764, p<0.001; *Significant*

Sequence length: F(1,5)= 12.233; p=0.017; *Significant*

Sequence length x electrode separation: F(6,30)=0.761; p=0.606; N.S.

*Repeat analysis, excluding extreme values for tone B (e19 and e21):*

Electrode separation: F(4,20)= 14.151, p<0.001; *Significant*

Sequence length: F(1,4)= 13.986, p=0.013; *Significant*

Sequence length x electrode separation: F(4,20)=1.038, p=0.412; N.S.

In the *apical* direction there is a large effect of electrode separation on thresholds delays, and a significant main effect of sequence length (greater thresholds delays for the triplets than for the full-length sequences); the two curves (for full-length sequences and for the triplets) are essentially parallel, with no interaction between electrode separation and sequence length. In the *basal* direction, electrode separation is again a significant effect, but the effect of sequence length did not quite reach significance. However, there is clearly a trend towards higher threshold delays for the triplets than in the full-length sequence condition. There is again no interaction between electrode separation and sequence length, although the slope of the curve for the full-length sequences appears somewhat steeper.

### 5.4.2 Individual data

Results for the 6 listeners who completed both task 1 (full length sequences) and task 2 (triplets) are shown in figure 5.3; results from the 2 listeners who only completed task 1 (full length sequences) are shown in figure 5.4. For the purposes of illustration, where threshold delays were greater than 40 ms, a symbol has been plotted at the 40-ms point on each graph, with an upward-pointing arrow. Similarly, where thresholds were found to be less than 10 ms (the shortest delay tested), a symbol has been inserted at the 10-ms point with a downward-pointing arrow.

Although some variation is apparent in individual listeners' results, some common patterns are observable. The strong dependence of threshold delay on electrode separation is obvious for both the full-length sequences and triplets data, with greater separations of tone B from electrode 11 in both the apical direction (higher electrode numbers) and basal direction (lower electrode numbers) generally associated with higher thresholds delays. Also, thresholds for the triplets were generally higher than those for the full-length sequences for most listeners.

Individuals showed some interesting differences in overall performance. For example, L2 showed thresholds above 40 ms (the highest delay used) on the full-length sequences for all the electrodes across the array except for the three centred on e11 (e10, e11, e12), and a very similar pattern for the triplets. In contrast, L11 had thresholds of 10 ms or better for the full-length sequences on all the electrodes from e3 through e14, with only slightly higher thresholds for the 4 most apical and one most basal electrode. However, for the triplets she showed generally higher thresholds with a clear dependence on electrode position (higher thresholds for greater separations from e11).

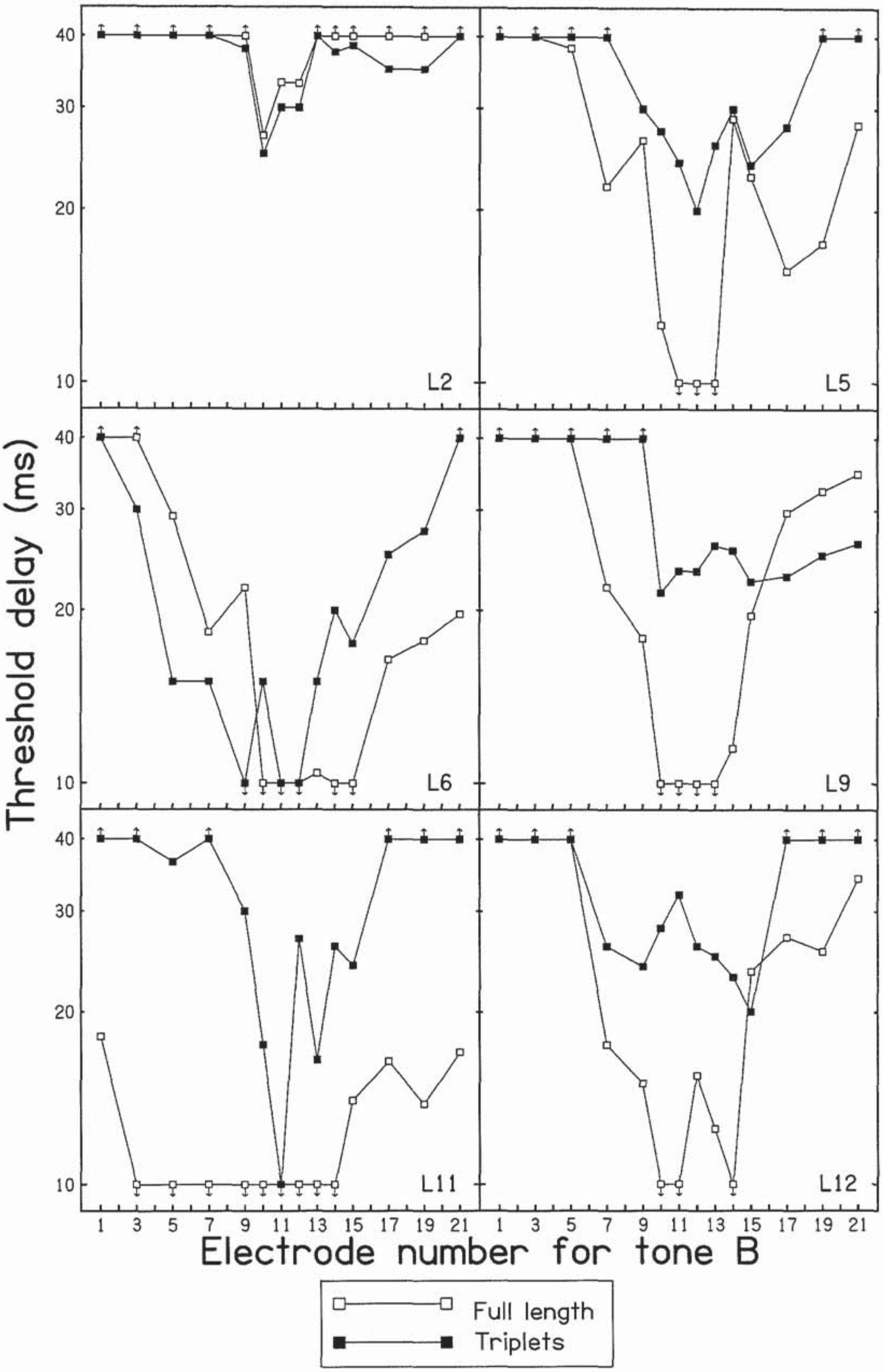


**Figure 5.3**. Individual listeners' results: 6 listeners who completed both task 1 (full length sequences) and task 2 (triplets).

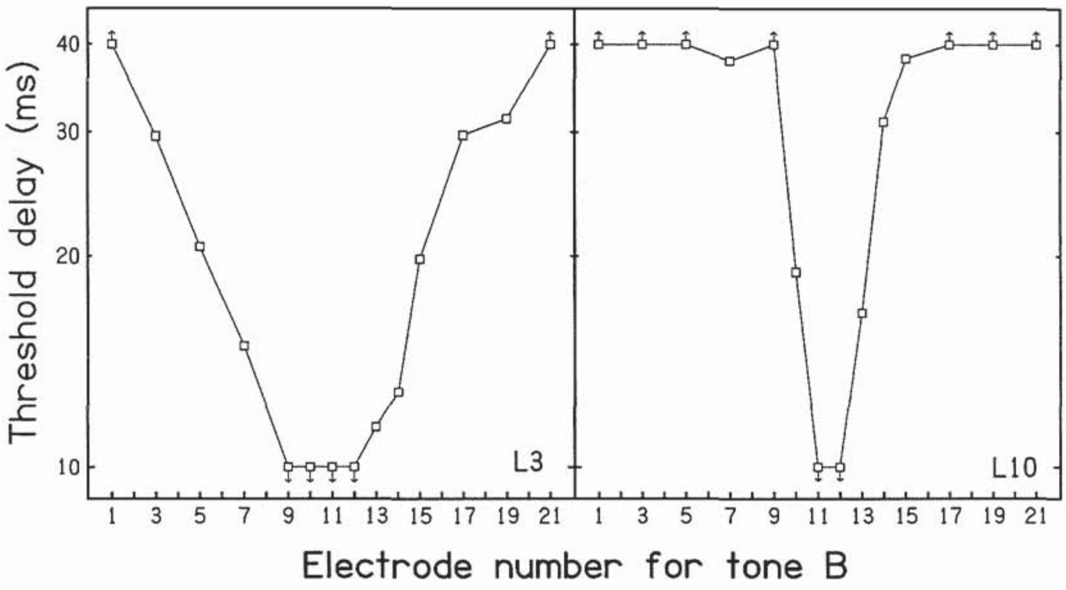


**Figure 5.4.** Individual results for the 2 listeners who completed task 1 (full length sequences) but who were unavailable to complete task 2.

### 5.5 Discussion

The main findings with implant listeners can be summarised as follows:

1. Increased electrode separations between tone A and tone B are associated in all listeners with worse performance (higher onset delay thresholds for tone B) for both the full-length sequences and triplets.
2. For most listeners, worse overall performance (higher onset delay thresholds) was generally obtained with the triplets than with the full-length sequences. This differs from the results reported by Hong and Turner (2006), where the opposite was found.
3. No interaction between sequence length and electrode separation was found.

The first of these findings is not surprising, and is consistent with other reports of temporal judgements with short sound sequences of different frequencies (e.g. van Noorden, 1975; Grose et al., 2001) or which stimulate different electrodes (e.g. Hanekom and Shannon, 1998). The results demonstrate that implant listeners found judgements about the relative timing of sounds more difficult when they stimulated more widely spatially separated electrodes, presumably resulting in greater perceptual differences, which were likely to be mainly differences in pitch or timbre.

The second finding, i.e. generally worse performance with the short triplets than with the 2.4-sec long sequences, is consistent with anecdotal reports from the listeners that they found task 2 (triplets) significantly more taxing than task 1. Whilst better performance in the triplets condition might have been expected as a result of a lack of build-up of segregation, this was more than offset by the small number of cycles available for listeners to detect the relative timing of tones A and B. Although this finding is consistent with the results reported by Hong and Turner (2006) for the condition where tone A=tone B (i.e. tone B was on electrode 11), they reported that the reverse was the case when the ratio between the frequencies of tone A and tone B was 1:1.5. The equivalent frequency separation in the present experiment would be a difference from e11 (centred on 1938 Hz) of 969 Hz (see table 5.1). This would be produced by positioning tone B on e8 (more basal/higher frequency than e11). No such reversal was consistently found in this experiment; although e8 was not used, results on electrodes adjacent to it were again worse for the 3-tone triplets than for the full-length sequences in the mean data and for all listeners other than L2. These results are therefore not consistent with those of Hong and Turner.

The third finding, that there was no evidence of an interaction between electrode separation and sequence length, also contrasts with the results of Hong and Turner (2006). None of the listeners tested here showed a crossover in their curves of threshold delay vs. electrode separation (aside from L9, for electrodes 13 to 21) as was reported by Hong and Turner. While it is not possible to prove a null hypothesis, it is worth noting that the analysis of variance had sufficient power to show significant effects for electrode separation and sequence length.

If the increased thresholds associated with greater electrode separations were providing a measure of increased automatic stream segregation, then a stronger dependence on electrode separation would have been expected for the full length sequences (which were of sufficient length to allow segregation to build up) than for the short triplets. The latter were too brief to have allowed any appreciable build-up to have occurred, and yet similar slopes for the electrode separation / thresholds delay functions were seen to those obtained with the full-length sequences. The results observed here do not support Hong and Turner's conclusions.

Two alternative explanations for these findings should be considered. First, for these listeners, the duration of the full-length sequences may have been insufficient to allow stream segregation to build up, though this seems unlikely. Second, that the expected effect is a small one and would require a greater number of listeners to obtain a significant finding.

One can conclude that the results do not provide any convincing evidence that stream segregation  occurred for these implant listeners on this task. This task was intended to provide an indirect measure of streaming; segregation into two streams should lead to worse performance, despite the efforts of the listeners. The results did show that increased electrode separations between successive tones led to worse performance on judgements about their relative timing, presumably as a result of perceptual differences between them (i.e. pitch differences). However, this effect was found for short sequences (tone triplets) to the same degree as for the longer sequences; it is therefore not possible to conclude that this finding was mediated by stream segregation.

In summary, the results of experiments 1 to 3, using either subjective or objective measures, have not provided convincing evidence that *involuntary* stream segregation takes place with cochlear implant listeners. In order to explore whether *voluntary* segregation, requiring attention, is possible for these listeners, a task that require listeners to *select* patterns of sounds from a background of distractor sounds provides an alternative approach, and this is employed in the next experiment.

# Chapter 6: Experiment 4

## 6.1 Introduction

Experiments 1 and 3 were concerned with stream segregation in cochlear implant (CI) listeners, using firstly a task requiring subjective reporting of segregation, and secondly a task where automatic segregation, if it occurred, would have impaired performance. The results so far have provided little or no evidence to indicate that *automatic* stream segregation makes a substantial contribution to the perceptual experience of most CI listeners. Rather, many of these listeners may have to rely mainly on *schema-based* selection to hear out a subset of acoustic elements from a sequence as a separate perceptual stream. Such a reliance on attentional mechanisms would inevitably limit the ability of CI listeners to cope with complex listening environments, particularly given that the sensory information they receive from their implants is impoverished.

Bregman (1990) proposed two different processes in the construction of auditory representations; firstly 'primitive' scene analysis and secondly, a schema- driven process. The first of these is driven by the incoming acoustic data (i.e. a 'bottom-up' process). The latter process is presumed to involve the 'activation of stored knowledge of familiar patterns or schemas in the acoustic environment and of a search for confirming stimulation in the auditory input' (Bregman, 1990; page 397). One example of a listening situation that would be expected to be dominated by such a process is when the listener has to select and recognise a familiar melody or pattern of tones from a sequence containing interleaved interfering or distractor sounds. An early example of an experiment using a task of this type was reported by Dowling (1973). He played 6 normal-hearing listeners temporally interleaved familiar melodies (e.g. 'Happy Birthday' etc), where the tones (square waves) of one melody alternated with the tones of another, presented in the 'rests' between notes. The listeners' task was to respond with the name of either of the melodies they recognised; all were familiar to them, and they had a list of the names of all 8 tunes used in front of them throughout the experiment. At first, the computer-generated melodies overlapped entirely in pitch range with each other; the frequency separation between the two melodies then gradually increased - on every 4th presentation one of the melodies was transposed up by a semitone. This continued until both melodies were correctly identified. He found that, in order for the melodies to be successfully identified, the pitch ranges of the melodies should not overlap.

As it was thought the results of Dowling's first experiment were confounded by the number of presentations, he conducted a second experiment using a short-term recognition memory paradigm. In this, listeners were presented with a standard 5-note melody followed by a comparison melody (both unfamiliar), which was temporally interleaved with a different 'background' melody. The task was to report whether the standard melody was present in the comparison sequence; the pitch range separation between the standard melody and the 'background' melody was either 0, 6 or 12 semitones. He found that, although performance was somewhat better with a pitch separation of 6 or 12 semitones, above-chance performance was possible even with zero pitch separation between the 'target' melody and the background melody. Dowling (1973) went on to demonstrate that if listeners were told the name of the melody to listen for, they were able to identify it successfully even when it was interleaved with a background melody in the same pitch range. Thus, active search for a familiar melody could overcome the effects of an interleaved distractor melody. This was a demonstration of a 'top-down', schema-driven process, where prior knowledge of what to listen for – combined with active listening – allowed for the perceptual separation of two streams. Automatic, primitive segregation on its own would not be expected to achieve this when both a target and distractor melody were within the same pitch range, unless some other perceptually salient difference between target and distractors was available as a cue for segregation. For example, Hartmann and Johnson (1991) found that differences between targets and distractors in timbre or location could improve the identification of pairs of interleaved familiar melodies.

A modification of Dowling's (1973) second task was used by Cusack and Roberts (2000). Their experiment used a trial structure containing 3 intervals. In the first was a target melody, comprising 5 pure tones. The second and third intervals contained a melody temporally interleaved with distractor tones. In one, the melody was the true target; in the other, the target underwent a 'melodic alteration' in which each tone was randomly shifted up or down in frequency by a certain amount to give a 'false target'. The listeners' task was to identify which interval contained the true target, and the size of the melodic alteration was varied in an adaptive procedure to determine a threshold value. This was measured for two different forms of contrast between the target tones and the distractors, which were either narrow-band noise bursts in the same frequency range as the target melody, or pure tones transposed upward by one octave. Threshold was also measured for pure tones in the same pitch range as a "no difference" control case. Distractors were identical in the second & third intervals on any given trial. Thus, a smaller melodic-alteration threshold was taken to indicate a greater degree of stream segregation between the target melody and distractors. The smallest threshold was found for the frequency difference condition, and the largest for a 'no difference' condition; the timbral difference gave thresholds about half-way between these two (see chapter 1 for more discussion). Cusack and Roberts (2000) went on to point out that, although the use of a schema to select sounds with a pure-tone timbre could have produced the observed results, it was not clear whether the differences in performance for the timbral contrast relative to the no-difference condition were due to *primitive* stream segregation or to *schema-based* selection.

Some additional evidence for the use of schema-based processes in a similar listening task was reported by Bey and McAdams (2002). Their task was a variation of that used by Cusack and Roberts (2000). Two unfamiliar six-tone melodies were presented successively to listeners who had to report whether the melodies were identical or different. One of the melodies was interleaved with distractor tones. In one condition, the comparison melody (without distractors) was presented after the mixed sequence; in another, the comparison melody was presented beforehand, giving the listeners a preview of what to listen for in the interleaved sequence. Performance was better in the latter condition compared with the first (a doubling in the d' for a mean frequency difference of 8 semitones), suggesting that prior knowledge of the melody to be extracted from the mixture enabled a schema-based process for segregation of target from distractors. This is consistent with the results of Dowling's (1973) third experiment, as discussed above. Thus, tasks requiring the recognition of interleaved melodies have been effective in experimental studies of auditory stream segregation. Such tasks probably involve both automatic and schema-driven segregation processes, and are probably dominated by the latter.

The interleaved melody task offers a potentially useful approach to the investigation of stream segregation in CI listeners (with the caveats outlined below), but thus far no published results from experiments of this type in CI listeners are available. However, the ability of CI listeners to recognise melodies has been evaluated as part of a wider interest in their appreciation of music. For example, Kong et al. (2004) used a closed-set melody-identification task with 6 CI listeners and 6 normal-hearing (NH) listeners. They pre-selected a set of 12 familiar songs that had been shown to give greater than 90% correct recognition in NH listeners; these were then presented on a musical synthesiser in two conditions, with or without rhythmic information (the latter contained only pitch information - the notes were of equal duration with equal-duration gaps between them). The NH listeners achieved near-perfect performance in melody identification in both the rhythm and no-rhythm conditions. In contrast, the CI listeners' performance was essentially at chance in the no-rhythm condition, and above chance but significantly poorer than for the NH listeners in the rhythm condition.

Kong et al. (2004) went on to evaluate the effect of the number of frequency bands in the signal on performance on the same task. They used a cochlear implant simulation program, in which the original broadband signal was divided into 1, 2, 4, 8, 16, 32 or 64 frequency bands in a noise vocoder. Chance performance in their task was around 8%. They found that the NH listeners achieved over 90% correct melody identification with only 4 frequency bands when rhythmic information was also available; however, the same level of performance was only achieved in the 'no-rhythm' condition with 32 frequency bands (16 bands gave around 70% correct identification). These results suggest that at least 16 or more functional spectral channels are needed to recognise familiar melodies without rhythmic cues; in reality CI listeners are likely to have no more than around 8 functional spectral channels (as discussed in chapter 2). If CI listeners were able to use all 8 of these functional channels for melody recognition as well as NH listeners, they would be expected to obtain up to 40% correct recognition in the absence of rhythm cues, based on Kong et al.'s NH listeners' results. Instead, they showed performance close to chance, similar to that which the NH listeners obtained with the implant simulation using only 1 or 2 frequency bands. Thus it would appear that on melody recognition CI listeners performed as if they had a single-channel implant, despite the fact that they had between 16 and 22 implanted electrodes (and are generally considered to have access to about 8 functional channels). Thus, even multi-channel CI users appear to rely heavily on rhythm information for melody recognition.

The design of any experiment based on interleaved melody recognition in CI listeners must take into account the findings of Kong et al. (2004); even good implant users who demonstrate high levels of open-set speech discrimination in quiet are likely to be very poor at melody recognition unless rhythmic cues are also available to them. Thus, in the absence of rhythmic cues, target melodies used in such tasks with CI listeners should be simple. Pitch differences between successive notes should be as salient as possible, bearing in mind the generally poor pitch perception CI listeners can achieve – the percepts conveyed by place of electrical stimulation via a cochlear implant may only be loosely defined as pitch. If discrimination between two melodies is required, the difference between them should be clear and obvious to the participants. These factors were considered carefully in the design of experiment 4, as described below.

Cochlear Implant listeners can hear pitch differences (or at least differences in timbral brightness) between different electrodes (e.g., as shown in experiment 2 of this thesis), so they might be expected to be able to select and recognise a simple melody from a background of distracting tones, provided that the distractors sound sufficiently different from the tones contained in the melody. This would provide some evidence that CI listeners can use *schema-based* selection to hear out a subset of acoustic elements from a sequence as a separate perceptual stream, even in the absence of automatic stream segregation. Directing interleaved distractor tones to electrodes spatially separated along the electrode array from the target/melody tones should provide pitch (or at least timbral brightness) differences that can be utilised to segregate melody from distractors. If CI listeners *are* able to make use of such cues, they should show better melody identification when the distractors stimulate a different and distinct part of the electrode array than when they overlap with the target melody tones. Also, performance should be better when a loudness difference cue is additionally available; i.e. the ability to ignore the distractor tones and so show better melody identification should be improved by attenuating the level of the distractor tones relative to the target tones. Note than any improvement seen when distractors are reduced in level cannot be accounted for by changes in masking, because distractor and target tones are presented at different times from one another.

Two main predictions were therefore tested in this experiment:

1. In a task requiring discrimination between two simple tone sequences (i.e., melodies) interleaved with distractor tones, performance should be better when the distractors stimulate a spatially distinct part of the electrode array from the melodies, compared to when the electrode ranges of the melodies and distractors overlap.
2. Better performance should also be achieved when the loudness of the distractors is reduced relative to that of the melody tones, compared to when they are at the same level.

## 6.2 Method

This experiment was a simple interleaved-melody identification task, in effect a simplified version of the task reported by Cusack and Roberts (2000). The task design was kept simple for the reasons discussed above. Two simple sequences of 5 pure tone ('melodies') were created, which were intended to be easily distinguishable from each other for most CI listeners. Other characteristics of the experimental set-up, including the configuration of the speech processor etc., were identical to those used in experiments 1 to 3 (see chapters 3 and 4 for full descriptions). Each 'melody' was centred around the middle of the electrode array; electrodes 9, 11 and 13 were stimulated. A spacing of two electrodes was used between successive sounds in an effort to provide a clear pitch contrast between them. The sequence of stimulation for melody 1 was on electrodes 9-11-13-11-9. This should lead to a falling pitch followed by a rise, as higher numbered electrodes are located towards the apical end of the electrode array and therefore should produce a lower pitch percept. For melody 2, the order was 13-11-9-11-13 (thus giving a rising pitch followed by a fall). Thus, the same set of electrodes was used in both melodies, and memorising and discriminating between them should be straightforward. No rhythm cues were available, as all tones were equal in duration and presented at the same rate.

The 'target' melodies were interleaved with distractor tones, selected randomly from five different electrode ranges as shown in table 6.1. Distractor sets 1 and 2 each stimulated electrodes at the far basal and apical ends of the array, respectively. Given that the melody tones were concentrated on the central portion of the array, there was a distinct spatial separation between these distractors and the melody tones. Distractor sets 3 and 4 stimulated ranges of electrodes *neighbouring* but *not overlapping* with the target melody tones (more basal and more apical, respectively). Finally, distractor set 5 used electrodes from the same range as the target melody tones (9-13).

As well as varying the location of the distractor tones, their level relative to the target melody tones was systematically varied. Without carefully loudness-balancing the stimuli relative to each other, it is not possible to be sure of their precise relative loudness. Instead, the loudness of the distractors was expressed in terms of percentage of dynamic range. The threshold (T) and maximum comfortable level (C) for each electrode was measured using standard clinical methods before running the experiment. The target melody tones were always presented at C-level. The stimulus level for the distractor tones was calculated from the dynamic range. For example, if the measured T level was 100 stimulus units and the C level was 200 stimulus units, the dynamic range would be 100 stimulus units (stimulus units are equivalent to log current level). The level of the distractor stimuli on each electrode in this example would be as follows:

100%DR (100% of dynamic range): 200 stimulus units
75%DR: 175 stimulus units
50%DR: 150 stimulus units
25%DR: 125 stimulus units

Following training in recognition of each melody in the absence of any distractors, the task required listeners to report which melody (1 or 2) was presented in the experimental interval, when the distractors were present.

### 6.2.1 Stimuli

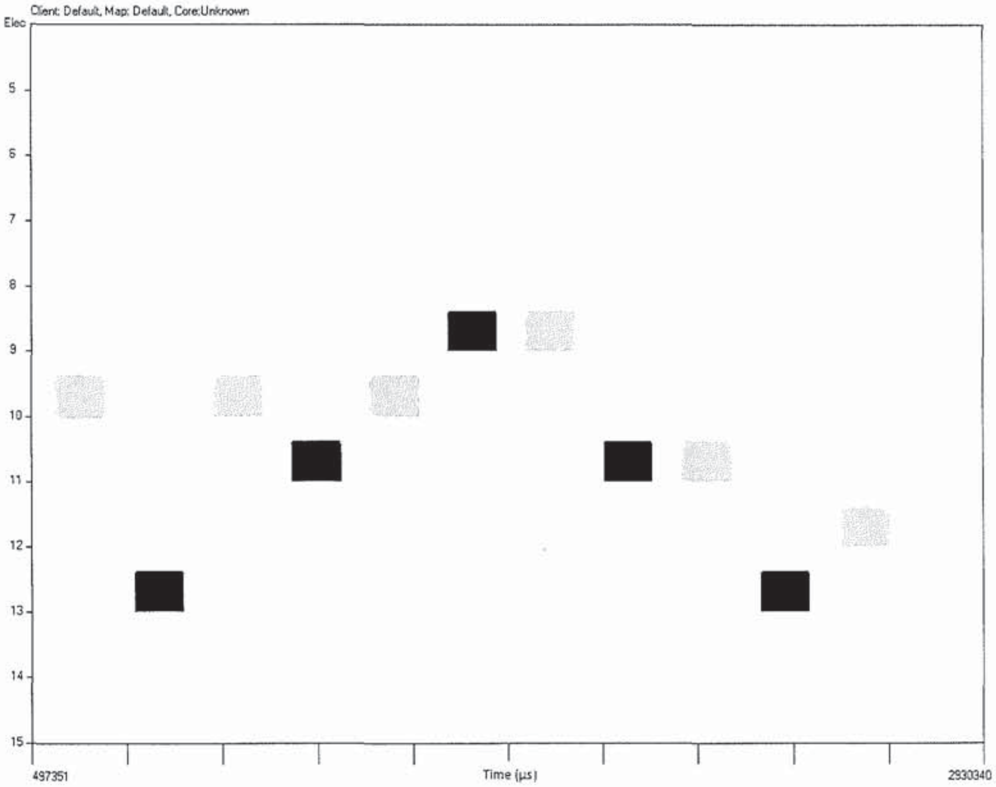
Individual stimuli for both the target melodies and distractors were 120-ms duration pure tones, with 10-ms rise and fall times. As in experiments 1 to 3, the pure tones used were at frequencies equal to the centre frequencies of the channel allocations defined in the experimental speech processor (see table 3.2, chapter 3), which was set up such that only one electrode at a time could be stimulated. As before, for all listeners, the pulse rate used in this study was 900 pulses per second and the pulse width used was 25 μs. Figure 6.1 shows the sequence of sounds contained in each of the melodies.

Either melody 1 or 2 was presented at random in the experimental interval (equal total number of each in each experimental block). The tones contained in each melody were temporally interleaved with 6 distractors, randomly picked from one of the ranges shown in table 6.1. In each case the resulting total tone sequence began and ended with a distractor tone as shown in figure 6.1. This ensured that listeners could not perform the task correctly simply by listening for the first or last note. There were short silent intervals (80 ms in duration) between the individual tones of the interleaved sequence.

**Table 6.1**. Details of electrode ranges in each distractor set

| Distractor set | Full name | Abbreviation | Electrode range |
|---|---|---|---|
| 1 | Distant/basal | D/B | 1-5 |
| 2 | Distant/apical | D/A | 18-22 |
| 3 | Neighbouring/basal | N/B | 4-8 |
| 4 | Neighbouring/apical | N/A | 14-18 |
| 5 | Overlapping | OV | 9-13 |

A


Client: Default, Map: Default, Core:Unknown
Elec
5
6
7
8
9
10
11
12
13
14
15
497351
Time (µs)
2930340

B

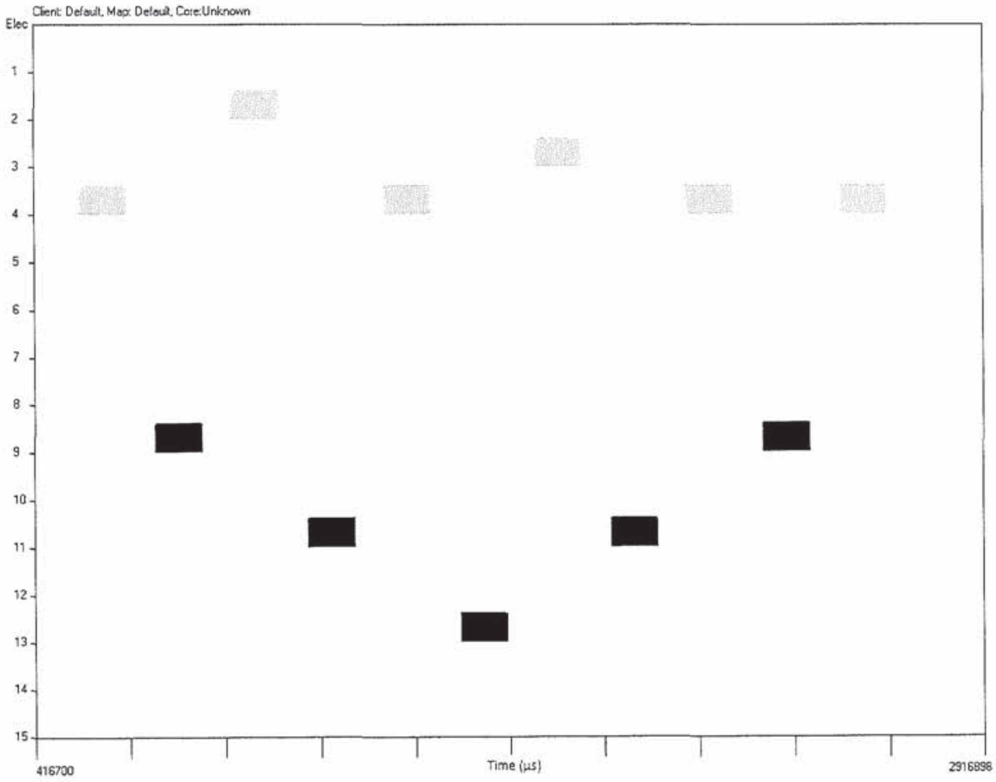

Client: Default, Map: Default, Core:Unknown
Elec
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
416700
Time (µs)
2916898

**Figure 6.1:**'Electrodograms' showing two examples of interleaved sequences. (A) Shows target melody 1 (rise/fall in pitch) interleaved with distractor tones in the overlapping condition (OV), i.e. taken at random from within the same electrode range as the target tones (9 to 13). (B) Shows target melody 2 (fall/rise in pitch) interleaved with distractor tones in the distant/basal condition (D/B), i.e. taken at random from the electrode range 1 to 5. The filled rectangles illustrate when electrical stimulation was present on each electrode. Black boxes show the target tones; grey boxes show the distractors. The ordinate shows electrode number and the abscissa shows time elapsed in µs. Only the range of electrode numbers that encompasses those stimulated in each case is illustrated.

### 6.2.2 Listeners

Listeners were 6 experienced adult cochlear implant users (4 female, 2 male) with the Nucleus CI24 device, all with full electrode insertions and normal electrode pitch ranking. Demographic and other details are shown in table 6.2. Three listeners (L1, L3 and L6) had also participated in experiment 1. All listeners were screened to ensure that they were able to reliably discriminate between the two target melodies when presented in isolation (without distractors). A criterion was set of obtaining 100% correct recognition on at least two successive experimental blocks (10 repetitions per block) following practice runs. Several potential participants screened in this way were rejected as they found the task too hard and were unable to perform above chance, even after training. Additionally, 3 normal-hearing listeners completed the task for comparison purposes. Their discrimination was screened in the same way and they had no difficulty in discriminating the two melodies.

**Table 6.2.** Demographic and clinical details of the 6 cochlear implant listeners. Those marked with an asterisk had not taken part in any previous experiments

| Listener | Age | Gender | Open set speech recognition score (BKB sentences in quiet, % correct) | Pulse rate (pulses per second) in normal daily use | Pulse width (µs) in normal daily use |
|---|---|---|---|---|---|
| L1 | 42 | F | 80 | 1200 | 25 |
| L3 | 61 | F | 91 | 1200 | 25 |
| L6 | 25 | M | 98 | 250 | 25 |
| L11 | 45 | M | 92 | 900 | 25 |
| L13* | 51 | F | 100 | 1200 | 25 |
| L14* | 25 | F | 90 | 250 | 25 |

### 6.2.3 Procedure

This experiment used a single-interval 2AFC design. Each interval contained randomly one or other of the target melodies. Following practice runs without any distractors present, the task was made progressively more difficult by introducing the distractors and increasing their level in each successive experimental block, in the sequence: 25% DR (Dynamic Range), 50% DR, 75% DR, and finally 100% DR. Each block contained 100 presentations, made up of 20 for each distractor range and presented in a random order; each melody occurred with equal likelihood (50% each) within each block. The 3 normal-hearing listeners were tested using target tones and distractor tones at equal levels (65 dB SPL), presented via headphones (Sennheiser HD535) in a sound-attenuating room. In this case, the frequencies of the tones used corresponded to those used for the cochlear implant listeners (see table 3.2, chapter 3) (N.B: although the same frequency tones were used for both NH and implant listeners, the resulting pitch percepts were likely to have differed due to the unpredictable frequency-to-place mapping present in cochlear implant listening).

### 6.2.4 Task and responses

The purpose of the experiment was explained to the listeners; they were instructed both verbally and in writing (see Appendix 4). All listeners received training and practice on the task prior to experimental runs. Care was taken to ensure that listeners understood the task, were able to respond appropriately, and that any task learning that was evident was essentially complete before the experimental runs began. Each listener was seated in front of a computer screen and keyboard. Before each presentation, the computer screen displayed the words 'Sequence 1 or 2?'. A visual representation of each melody was provided as a reminder of what to listen for. This showed the expected pitch changes for each melody. They were instructed to press either the '1' or '2' key on the keyboard after each trial to indicate which target melody they had heard. Feedback was provided on the screen following each response, displaying 'Correct' or 'Incorrect' as appropriate. The experimental software recorded each response and whether or not it was correct.

## 6.3 Results

Responses in each condition were analysed in terms of percentage correct responses. Mean results for all 6 CI listeners are shown in figure 6.2. There is a clear and progressive improvement in performance with decreasing distractor level. When averaged across distractor set, mean performance for distractor levels at 100%, 75%, 50%, and 25% of dynamic range was 55%, 62%, 92%, and 96% correct, respectively. Mean scores were close to chance for the 100% DR condition, slightly better for the 75% DR condition, and were generally above 90% correct for the 50% DR and 25% DR conditions.

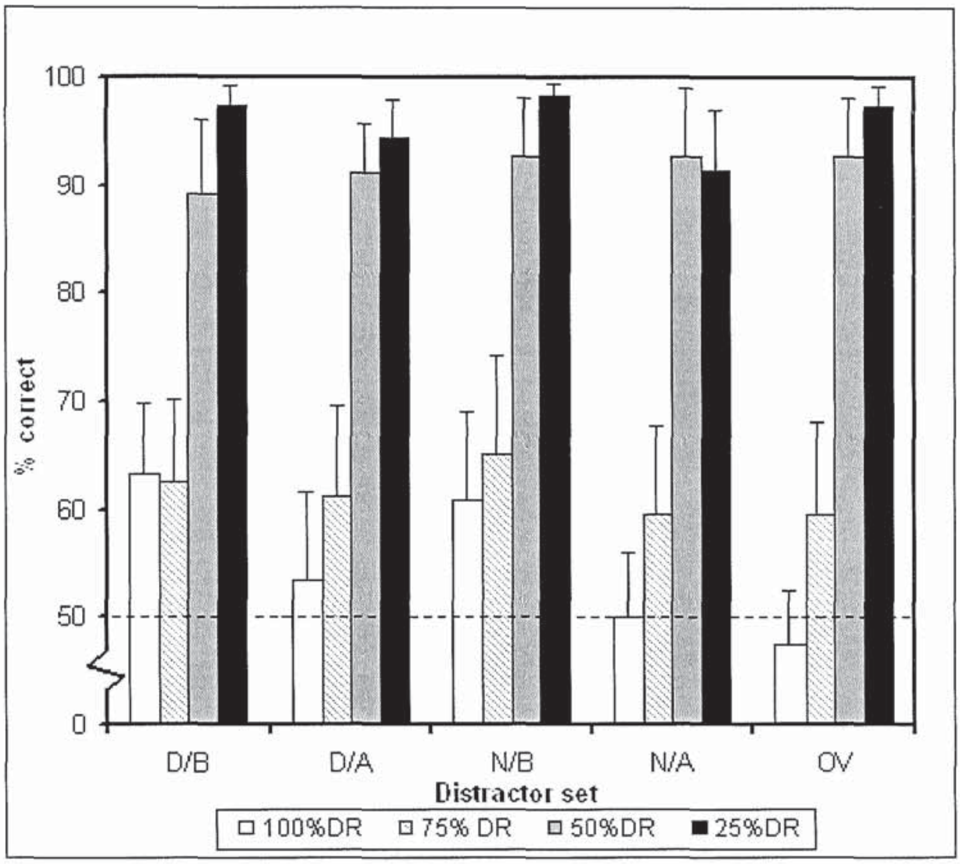


**Figure 6.2**. Mean results for all 6 cochlear implant listeners. Clustered columns show results for each distractor set. Standard errors for each mean are shown by vertical lines. Abbreviations for the different distractors sets are as described above. The four different levels of distractors are shown by the differently shaded columns. Results for the 'no distractors' condition are not shown: all listeners obtained 100% correct performance in that condition. Horizontal dashed line indicates a chance level score.

Contrary to expectation, which distractor set was used appeared to have little effect on performance. A within-subjects, repeated-measures ANOVA showed a highly significant effect of distractor level: F(3,15)=16.56, p<0.001, but there was no significant effect of distractor set: F (4,20)=1.29, p=0.307, and no significant interaction between distractor set and level: F(12,60)=1.15, p=0.338. The ANOVA was repeated using arcsine-transformed data, with the intention of reducing the heterogeneity of the data. This again showed a highly significant effect of distractor level: F(3,15)=16.531; p<0.001, no significant effect of distractor set: F(4,20)=1.582, p=0.218, and no significant interaction between distractor set and level: F(12,60)= 1.211, p=0.297. Pairwise comparisons between means for distractor levels are shown in table 6.3. Four out of six comparisons between mean scores for different distractor levels showed a significant difference at p<0.05; there was no significant difference between 100% DR and 75% DR, or between 50% DR and 25% DR.

**Table 6.3**. Pairwise comparisons between means for distractor levels (averaged across distractor sets).

| Distractor levels compared | Mean difference in % correct | Significance level |
|---|---|---|
| 100% DR vs 75% DR | 6.64 | 0.472 |
| 100% DR vs 50% DR | 36.74 | **0.004** |
| 100% DR vs 25% DR | 40.80 | **0.001** |
| 75% DR vs 50% DR | 30.10 | **0.023** |
| 75% DR vs 25% DR | 34.16 | **0.005** |
| 50% DR vs. 25% DR | 4.06 | 0.053 |

Figure 6.3 shows individual results for the 6 CI listeners. Overall performance on this task in the 100% DR condition (distractors at equal loudness levels to the target tones) was generally close to chance, with no systematic difference between the different distractor sets. One exception was L13, whose performance differed from the other listeners; she showed the best scores for the two basally located distractor sets D/B (distant / basal) and N/B (neighbouring / basal), both well above chance (90% and 95% correct

respectively) and, surprisingly, better than in the 75% DR condition. In the 75% DR condition (distractors at 75% of dynamic range), all the CI listeners aside from one (L11) again showed performance close to chance; in contrast, L11 showed strikingly good performance, with 100% correct scores for three of the distractor sets and 95% correct for the other two (including the overlapping set). The best performance was seen for the two lowest levels of distractor, i.e. at 50% and 25 % of dynamic range.

Results for the 3 NH listeners are shown in figure 6.4. All results are for distractors of equal level (dB SPL) to the melody tones (equivalent to the 100% DR condition for the CI listeners). A within-subjects ANOVA showed a significant effect of distractor set: F(4,8)= 12.298, p=0.002. All three performed well above chance level (two with near-perfect performance) for the four conditions in which distractors were in a different frequency range to the target tones. In the OV (overlapping frequency range) condition, two listeners gave responses close to chance, while one (N2) achieved a score of 85% correct in this most challenging condition. Given that there were only two possible target melodies, the more successful listener may have used a particular listening strategy, such as attending to the pitch of the second note (which is the first target tone) and identifying the melody correctly based only on that pitch. The provision of feedback would have helped the learning of such associations.

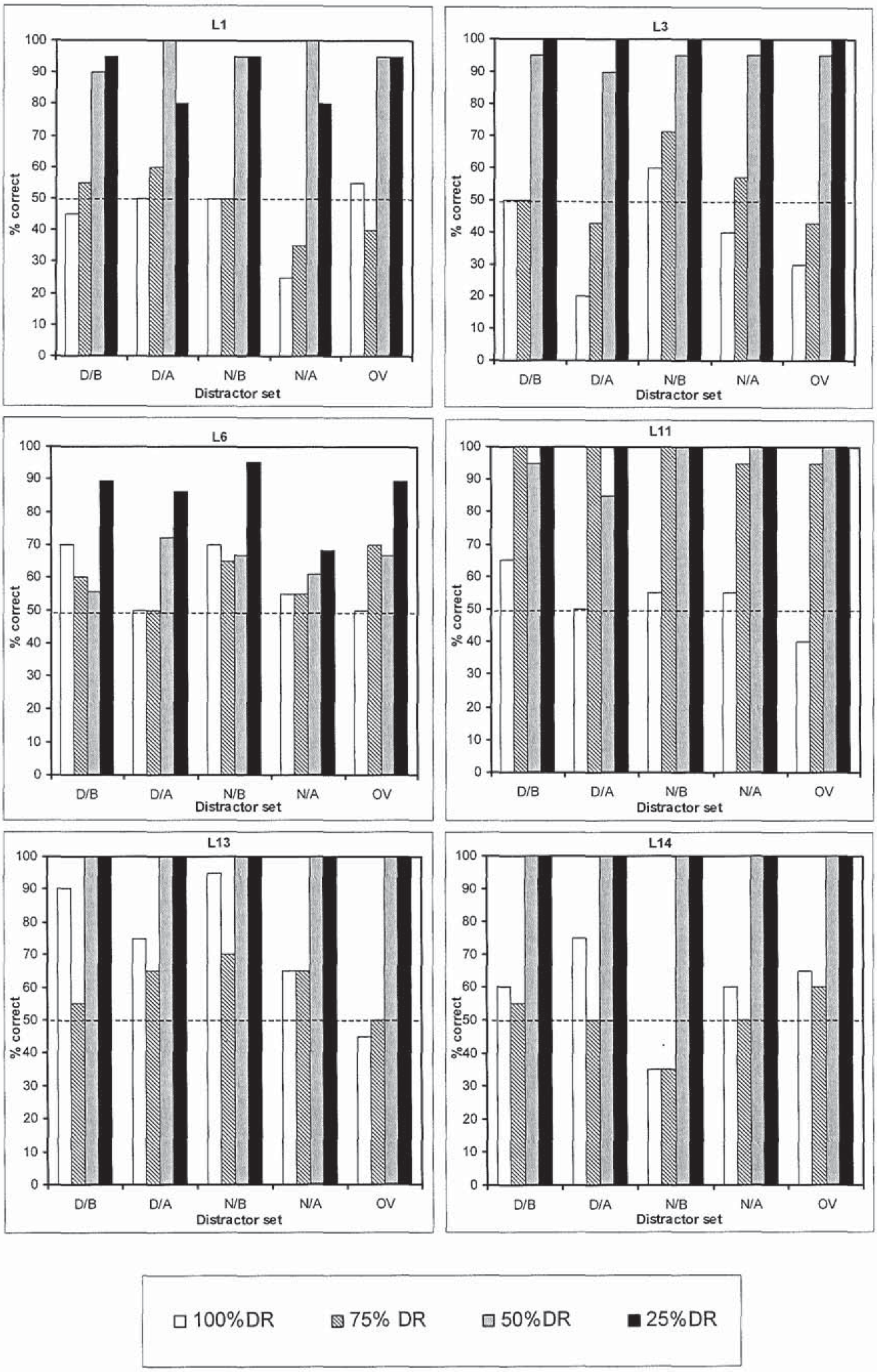


**Figure 6.3.** Individual results for the 6 CI listeners. Clustered columns show results for each distractor set. Abbreviations for the different distractors sets

are as described above (see table 6.1). The four different levels of distractors are shown by the differently shaded columns. Results for the 'no distractors' condition are not shown: all listeners obtained 100% correct performance in that condition. Horizontal dashed lines indicate a chance level score.

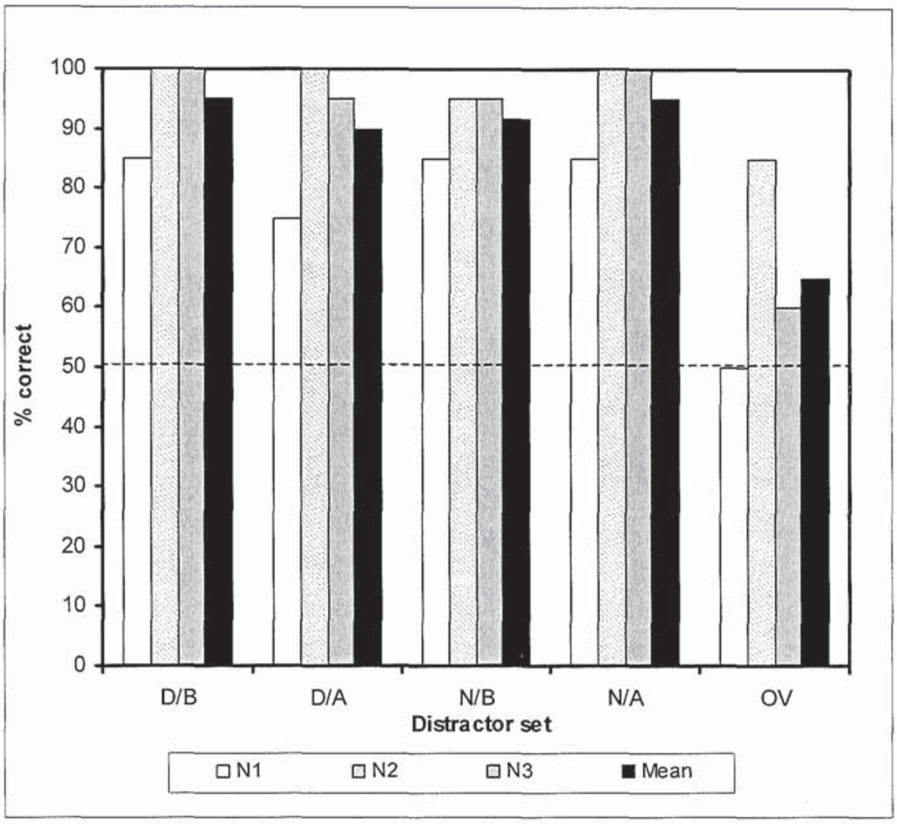


**Figure 6.4**. Individual results for the 3 normal-hearing listeners. NB: distractors were at equal dB SPL to the target tones (i.e. equivalent to the 100% DR condition for the CI listeners). Results for individual listeners and the mean data are represented by the differently shaded columns. Clustered columns show results for each distractor set.

## 6.4 Discussion

With some exceptions, individual performance by these CI listeners on this task was reasonably consistent. Aside from one listener, whose results differed from those of the rest of the group, they were not able to reliably identify the target melodies when they were interleaved with distractors at equal loudness to the target tones (i.e. 100% of the dynamic range) and responded at or near chance, irrespective of the electrode range occupied by the distractors. With distractors attenuated to 75% of the dynamic range, performance remained only slightly above chance, again aside from one (different) listener. At 50% of dynamic range, 5 out of 6 were able to demonstrate near-perfect performance, and when the distractors were reduced to only 25% of the dynamic range (close to threshold), all showed above chance (and usually near-perfect) performance.

The fact that the mean results for each distractor set did not differ significantly from each other shows that these CI listeners were generally (with the exception of L13) not able to benefit from any pitch (or other) differences between target tones and distractors when listening for the target melody and trying to ignore the distractors. In principle, it is possible that the inclusion of intermediate distractor levels (between 50% and 75%) might have better distinguished between different distractor sets. However, even if that had been true, a much smaller effect of distractor set would be implied than was originally anticipated. In contrast, the 3 NH listeners were all able to score well above chance so long as the distractors were drawn from a different frequency range to the target melody tones - even though the distractors were of equal loudness to the target tones.

Thus, the performance of these CI listeners on this task differs substantially from those obtained with NH listeners; of the two predictions given in the introduction, only the second has been upheld. Based on these results, it would seem that CI listeners are generally *not* able to utilise differences in pitch or timbral brightness between target sounds and distractors resulting from stimulation of different sections of the electrode array, to segregate them and attend to the target sounds while ignoring the distractors. However, they *are* apparently able to ignore the distractors and successfully attend to the target melodies, provided that the distractors are reduced in level to 50% or less of the dynamic range on each electrode. This would be equivalent to attending to sounds at the maximum comfortable level, while ignoring distracting sounds that are clearly audible, but softer.

The fact that stimulation of electrodes in differing portions of the electrode array did not enable listeners to ignore distractors and attend to the target tones suggests that that the perceptual differences between sensations evoked by stimulating different places along the electrode array may not be great enough to support schema-based selection. This may be consistent with the results of experiments 1 and 2, where no convincing evidence of stream segregation based on electrode separation was found. However, CI listeners *can* perceive pitch differences between different electrodes to varying degrees, as shown in experiment 2, and were able to reliably discriminate between two simple melodies in the absence of any distractors. In view of this, one might have expected the benefits of differences in electrode ranges and in level to have combined to enable schema-based selection of the target melody from the distractors, but this was not found. This may not be entirely surprising, in the light of the findings of Kong et al. (2004) discussed above; CI listeners are very poor at melody recognition and appear to rely on rhythm cues when listening to music. In effect, their performance on melody recognition in the absence of rhythm cues is similar to that of a single-electrode implant; this would make segregation of a target melody from interleaved distractors very difficult or impossible, although in the absence of distractors these listeners were able to discriminate between the two target melodies with 100% accuracy.

# Chapter 7: Experiment 5

## 7.1 Introduction

Experiments 1 to 4 evaluated the role of sequential stream segregation in cochlear implant listeners; overall, the results have not provided convincing evidence of stream segregation. However, the ability of implant listeners to perceptually segregate simultaneous sounds (concurrent sound segregation) has not so far been examined in these listeners. In real life, concurrent mixtures of sounds have to be analysed by our auditory system in many common listening situations, for example when following speech in the presence of competing speech, or when listening to polyphonic music. Therefore, it is important to consider the ability of implant listeners to segregate sounds in a simultaneous grouping task.

### 7.1.1 Concurrent sound segregation in normally hearing listeners

In normal-hearing (NH) listeners, a number of cues are used to allow segregation of concurrent sounds in everyday listening. For example, when separating target speech from competing speech sources, differences in F0 and intonation between talkers can be exploited to improve intelligibility (Brokx and Nooteboom,1982; Bird and Darwin, 1998). Differences in F0 can also be used in the identification of simultaneous vowels, and this benefit is reduced in hearing impaired listeners (e.g. Arehart et al., 1997).

Another important cue for the segregation of concurrent sounds is onset asynchrony; the auditory system makes an assumption that sounds that start and finish simultaneously are likely to have originated from the same source, and so are perceptually grouped together. Conversely, a component that starts at a different time to a set of other components is likely to be segregated perceptually. This makes sense, as it is unlikely that acoustic elements arising from different sources will start and stop at the same time. For example, Darwin (1984) showed that a harmonic that starts before the rest of a vowel sound makes less contribution to vowel quality than one that starts at the same time as the rest of the harmonics in the vowel. Similarly, Roberts and Moore (1991) showed that the shift in phoneme boundary resulting from the addition of a pair of extraneous tones to each of a series of vowels declines in conditions of onset asynchrony (i.e. when the onset of the tones preceded that of the vowels). Also, Darwin and Ciocca (1992) showed that when a slightly mistuned component in a harmonic complex tone started 160 ms or more before the remaining harmonics but stopped simultaneously with them, it made a reduced contribution to the pitch of the complex; it made no contribution if it started more than 300 ms before.

Harmonic relations provide an important cue that can be used in normal hearing for concurrent sound segregation (see e.g. Darwin, 2005). When spectral components have frequencies that are in a perfect harmonic relationship, those components tend to be integrated by the auditory system into a single perceived entity (Hartmann et al., 1990). For example, Moore et al. (1986) and Hartmann et al. (1990) have shown that if one component in a harmonic complex is mistuned by a sufficient amount, it is heard as a separate tone, standing out from the rest of the complex.

Another way in which a single pure-tone component can be caused to perceptually 'pop out' from a complex tone is if its intensity is momentarily reduced and then returned to its original level. This was demonstrated by Kubovy and Daniel (1983); they presented listeners with a 'chord' made up of 8 pure tones. When all the components in the chord had the same intensity, none of them dominated the perceived pitch of the chord. When each of the tones was reduced in level in turn for 80 ms and then abruptly restored to its former level, the pitch of that tone became dominant, allowing a scale or melody to be played. They showed that this was robust effect, which still could be perceived even for amplitude notches as shallow as 3 dB.

Kubovy and Daniel referred to this as segregation by momentary amplitude disparity; although there had been no overall increase in energy in the tone, a sudden rise in its intensity led to its perceptual dominance. Aspects of this phenomenon, relating only to *incrementing* target components, were studied in more detail by Bregman et al. (1994a,b). Their first experiment evaluated the effects of both abruptness and synchrony of onsets in the perception of clusters of pure tones. Listeners judged the order of two component tones, at 750 Hz and 850 Hz, in a 4-tone cluster which always began and ended with an 800 Hz tone. The velocity of the rise in intensity of the tones was varied, as well as the onset asynchrony; all 4 tones terminated simultaneously. They converted the resulting responses into a discrimination score, and found a strong effect of onset velocity on discrimination - i.e. more sudden onsets were associated with better discrimination of the order of the tones. Also, as expected, larger onset asynchronies gave better discrimination scores. In a

2nd experiment, they examined the role of sudden *offsets* in the same task, by temporally reversing the stimuli. Thus, the tone clusters had synchronous *onsets* but asynchronous offsets. Performance (as measured by discrimination scores which ranged from +7 (perfect discrimination) to -7 (calling all identical pairs 'different' and vice-versa)) was worse overall than seen in the 1st experiment; however, there was still a small effect of the abruptness of offset, i.e. more abrupt offsets helped a little in discriminating the order in which the tones ended. For example, the mean discrimination score for an offset time of 640 ms was 0.1, and for an offset time of 160 ms was 0.8). By comparison, the same difference in *onset* time in the first experiment gave an improvement in discrimination score from 0.7 to 2.8. Thus, the effect of *onset* was considerably larger than that for *offsets*.

Bregman et al. (1994a) referred to this beneficial effect of abrupt onsets (and to a lesser extent offsets) as a 'sudden-change' effect, and hypothesised that this represented a 'resetting' of the pitch-computation system. This is also consistent with real-life listening situations, where abrupt increases in sound mixtures tend to signal the introduction of new sources, while gradual changes are more likely to be caused by changes in on-going sounds. Many natural sounds, in particular percussive sounds, have sharp onsets but gradual offsets. However, Bregman et al. (1994b) pointed out a difficulty with their previous interpretation of the experiment described above. The amplitudes of the tones in the cluster they used had linear rising envelopes and a fixed stimulus onset asynchrony. As a result, there was a larger difference in the relative amplitude of the first two tones between the rapidly rising compared with the slowly rising clusters. This might have been responsible for the greater clarity of pitches in the rapid-onset cluster, which had led to the observation of better discrimination with more rapid onsets. To try and eliminate this problem, Bregman et al. (1994b) designed two new experiments. In the first, they presented two pure tones in succession at either 1000 Hz, 1500 Hz or 2000 Hz, in the middle of a 3.5-s burst of noise (band-pass filtered with upper and lower cut-off frequencies of 2500 Hz and 500 Hz); signal-to-noise ratios were 1, 2 or 3 dB. The rise-times of the tones were 30, 90, 270, 730, 910 or 970 ms, with total duration = 1 s. As they had linear rise and fall times and no plateaus, the last three envelopes were 'mirror images' of the first three (i.e. a tone with a 30 ms rise time had a fall time of 970 ms, and vice-versa). Listeners reported the order of the two tones by stating whether the pitch pattern had gone up or down. The authors had hypothesised that more rapid rise-times of the tones would lead to more accurate discrimination, as a result of processing being attracted to the frequency band of the tones more strongly for rapid onsets than for slower onsets. However, they found only a small decline in discrimination with longer rise times, and a much smaller effect than that seen in the previous experiment with clusters of tones.

The crucial difference between this experiment and the previous one was the fact that it concerned the detection of a signal masked by noise, as distinct from perceptual segregation of tones from a context of other tones. Bregman et al.'s (1994b) second experiment therefore used a complex tone consisting of a fundamental of 500 Hz along with its next 4 harmonics (i.e. 1000 Hz, 1500 Hz, 2000 Hz and 2500 Hz). The amplitudes of an adjacent pair of two of the 'inner' tones (1000 Hz, 1500 Hz, or 2000 Hz) were successively incremented, while the rest of the tones in the complex remained at a steady state. This led to the pitch of the incremented tones becoming salient and standing out from the rest of the complex. As before, the rise times for the triangular envelope shapes for the tones ranged between 30 ms and 970 ms; the sizes of the increments were 1, 3 or 6 dB. They found a strong effect of rise time, with the best discrimination scores obtained with the most sudden onset (30 ms rise time, discrimination score = 3.89 for a 6 dB increment) and declining with increasing rise times up to the maximum of 970 ms (discrimination score = 2.37, again for a 6 dB increment). There was also a small but significant effect of increment size; for example, for a rise time of 270 ms, a 1 dB increment gave a discrimination score of 2.54, and a 6 dB increment gave a discrimination score of 3.24. The explanation given by the authors was that the same 'sudden change' effect, that had led to the results seen in their previous experiment, also applied to increments in amplitude in otherwise steady-state tones in a complex. Thus, the increments activated the same sort of onset responses in the auditory system that are seen for true onsets. Their hypothesis was that this onset response does two things; firstly, it supplies frequency information enabling identification of the tone and secondly it 'points to' a narrow spectral region for enhanced analysis by other processes, enabling auditory scene analysis. A very weak or non-existent onset response was seen in their first experiment with a background of noise.

### 7.1.2 Concurrent sound segregation in cochlear implant listeners

Some of the cues for concurrent segregation that have been discussed in relation to normally hearing listeners cannot be expected to be available to cochlear implant listeners. In particular, cues based on harmonic relations are unlikely to be useful as individual harmonics are not effectively conveyed by any currently available combinations of electrode arrays and speech processing strategies; individual harmonics are generally not resolved, and the log spacing of implants channels does not provide any means of signalling harmonic relations between components. Also, as F0 is poorly transmitted in cochlear implant listening, it is unlikely that CI users will be able to make use of this as a cue for concurrent sound segregation. Nelson et al. (2002) showed that, unlike NH listeners, CI listeners are unable to take advantage of gender mismatches between target and masker speech when listening in a 'cocktail party' type situation. Nelson et al. (2003) showed that CI listeners are not able to obtain release from masking from modulated noise, while NH listeners can (particularly at an 8 Hz masker modulation frequency). However, some cues may in theory be available to CI listeners and might therefore be demonstrable in an experimental setting. One of these is onset asynchrony, which is commonly encountered in everyday listening situations and has been shown to be a robust cue in NH listeners (as discussed above). Another (that is more specific to cochlear implant listening) is differences in pulse rate. Implant listeners are able to discriminate pulse rate on single electrodes up to a frequency limit of around 300 Hz. Therefore, it would be reasonable to expect that this cue could be used for segregation of concurrent sounds. However, Deeks and Carlyon (2004) suggested, based on the results of their study using simulations of cochlear implant listening, that differences in pulse rate were unlikely to be useful for this purpose. They found that although speech recognition performance could be improved by presenting target speech at a higher pulse rate (140 pps) than masker material (80 pps), compared with when they were both presented at the same rate (both at 140 pps), performance was *worse* when the pulse rates for target and masker were reversed. They also found that presenting the target speech on only the odd-numbered channels of a 6-channel simulation and the masker speech only on the even-numbered channels (or vice versa) gave worse performance than presenting them both on all channels

A recent study by Carlyon et al. (2007) used a two-interval forced choice procedure to evaluate the effects of pulse-rate differences and stimulus-onset differences in cochlear implant (CI) listeners. In their task, 5 CI listeners (with the Nucleus CI24 device) were presented with a 400-ms duration 'mixture' created by simultaneous stimulation of 4 electrodes, each separated by 4 electrodes from its neighbour, distributed across the electrode array (the actual electrodes used varied between listeners, but for each listener the same set of 4 electrodes was used throughout). Stimulation on 3 out of the 4 electrodes in the mixture was at a fixed percentage of the listener's dynamic range (DR); 1 'target' electrode (one or other of the two inner electrodes) was stimulated at a higher percentage of the DR. Each mixture was preceded, 200 ms earlier, by a 400-ms probe on one of the four electrodes. One each trial, one interval contained a 'mixture' preceded by a probe identical to the pulse train on the target electrode in the mixture. In the other, the probe was the same as the pulse train applied to the other possible target electrode. Listeners had to identify which interval contained the mixture in which the probe was most clearly present (the mixture was actually identical in both intervals). In a baseline condition, only the increased level on the target electrode distinguished it from the other 3 sounds in the mixture. Three other conditions evaluated the beneficial effects of: (a) a difference in pulse rate on the target electrode compared to the other electrodes in the mixture (77 pps compared with 100 pps); (b) an onset delay of 200 ms applied to the pulse train on the target electrode and to the preceding probe (N.B: stimulation on all electrodes ended at the same time); and (c) an 'asynch' condition where the pulses on the target electrode were delayed by 5 ms relative to the other electrodes in the mixture. If any of these manipulations helped to perceptually segregate the target electrode from the rest of the mixture, performance should be better in that condition as the probe should be more salient in the 4-electrode mixture. No feedback was provided to listeners during the test procedure. Carlyon et al. found a small but significant beneficial effect of onset delay on performance, although the actual improvement (expressed in higher percent correct score) compared to the baseline condition was modest and less than 10 percentage points in 4 out of the 5 listeners (N.B: listeners were able to perform above chance in the baseline condition because of the increased level applied to the target electrode; for example, one listener's score in the baseline condition was around 88% correct, and increased to around 95% correct in the onset-delay condition). There was no significant beneficial effect of pulse-rate difference or of pulse asynchrony.

Two listeners took part in a further experiment in which no increased stimulus level was applied to the target electrode, so that in the absence of any additional cue, performance would be expected to be at chance level (50% correct). Both again showed performance above chance in the onset-delay condition, although for one listener the effect was quite small; their score was just under 60% correct). In a second experiment using an analogous task with NH listeners, Carlyon et al. (2007) again found a small benefit of a 200-ms onset delay. In this case, instead of just 4 stimuli quite well separated in frequency, as used with the CI listeners, the stimuli used were a mixture of harmonics 5 to 40 of a F0 of 100 Hz; the onsets of harmonics 13-17 or 26-30 were delayed by 200 ms in a 400-ms duration stimulus. They also showed that introduction of a spectral gap between target and non-target bands greatly enhanced performance. Based on this, they hypothesised that overlap of electrical stimulation caused by spread of excitation in the CI listeners had reduced the benefit of the stimulus onset delay in concurrent sound segregation. Therefore, improvements in channel selectivity would be expected to improve concurrent (as well as sequential) segregation.

## 7.2 Overview of experiment 5

As discussed, a number of different cues to concurrent sound segregation have been evaluated in NH listeners. The Carlyon et al. (2007) study showed only that stimulus onset delay (i.e. a difference in the temporal onset of one component in a complex sound) can provide a small benefit to concurrent segregation in CI listeners. The 'sudden change' effect described by Bregman et al. (1994b) for NH listeners, has not yet been explored in CI listeners; the aim of this experiment was to do so.

Two potential cues for segregation of a sound from a mixture of other sounds were evaluated. Firstly, a temporary increase in energy in the stimulation applied to one electrode, in the context of steady-state stimulation on others in a 'complex tone'. If implant listeners are able to detect such a change in energy, this should provide a 'pop out' cue for the electrode to which it is applied and elicit an audible pitch, as demonstrated in Bregman et al.'s (1994b) study. If CI listeners perform in a similar way to NH listeners, an abrupt rise in energy should provide a stronger segregation cue than a slow rise when the total energy contained in each increment is identical. Secondly, a temporary decrement in the stimulation applied to one electrode in a mixture, followed by a return to its previous value, should produce a sudden-change effect, similar to the experiment described by Kubovy and Daniel (1983). If CI listeners are able to use a sudden-change cue in this context, a pitch should again be elicited related to the electrode whose level has been manipulated, and this should be most effective when the *offset* (i.e. the rise to normal level at the end of the decrement) is more abrupt.

## 7.3 Method

The experimental approach used was very similar to that described by Bregman et al. (1994b). The main difference was that instead of a complex tone made up of a fundamental and 4 harmonics, a 'complex tone' created by stimulation of 5 electrodes evenly spaced across nearly the whole electrode array was used. Thus, the stimuli were more akin to a log-spaced set of components in NH listeners, in contrast to the harmonic complex used by Bregman et al. (1994b). Pure tones at the frequencies they used (500, 1000, 1500, 2000 and 2500 Hz) would have led to stimulation on electrodes 20, 16, 13, 11 and 9 respectively, if the same channel frequency allocations were used as for experiments 1 to 4. In view of the generally limited spectral resolution and problem of current spread known to occur with CI listeners, it was considered better to optimise the chances of resolution of the individual components in the complex by using widely and equally spaced electrodes.

The other difference between this experiment and Bregman et al.'s is the size of the increments and decrements used. They found that NH listeners could detect increments with abrupt onsets as small as 1 dB. However, in order to ensure that the task was manageable for CI listeners, larger increment sizes were tested, expressed in terms of percentage of dynamic range instead of in dB; the largest increment/decrement used was 50% of the dynamic range. This is roughly equivalent to around 60 dB in a NH listener with a dynamic range of 120 dB, and so might reasonably be expected to provide a clear cue for segregation.

Specific manipulations required careful control of stimulus parameters, and this was achieved by use of the 'Impress' software (developed at the University of Melbourne, in collaboration with the Cambridge MRC-CBU and others). This allowed creation and presentation of the combinations of stimulation on multiple electrodes required by the experimental design.

Stimulus presentation was via the 'Spear' experimental speech processor and patient response box. Listeners' own speech processors were not used.

Part one of the experiment employed *increments* in stimulus level, and part two employed *decrements*, in a 'complex tone' made up of stimulation on 5 well separated electrodes (3, 7, 11, 15, 19). The 'target' electrodes to which the increments or decrements were applied were always an adjacent pair within the interior set of three in the complex. The changes were applied to two electrodes in succession, in a background of steady-state stimulation on 3 other electrodes evenly spaced across the electrode array for the Nucleus CI24 cochlear implant (see figure 7.2). The 'steady state' stimulation was set at 50% of the dynamic range for each electrode throughout. This should produce a medium-loud sensation level. Increment sizes were an additional +12.5%, +25% or +50% of the dynamic range; the last of these resulted in an increase in level peaking momentarily at C level (maximum comfortable level). Decrements were at -12.5%, -25%, or -50% of dynamic range; the last of these resulted in a decrease in level down to a momentary minimum at T level (threshold). Threshold (T) and maximum comfortable levels (C) for each of the electrodes to be stimulated were measured using a standard psychophysical procedure prior to the experimental runs. Increment and decrement sizes were then calculated for each listener and each electrode; an example is shown in table 7.2. A constant pulse rate of 900 pps and pulse width of 25 µs was used throughout for all listeners. Increments/decrements were applied only to electrodes 7 (H), 11 (M) or 15 (L). The H, M and L notations were used to indicate the expected pitch differences between the electrodes concerned;

for example the location of e7 is more basal compared to that of e11 and so should produce the highest pitch (or brightest timbre) out of the three, and so was labelled H. Stimulation remained at 50% of dynamic range throughout on the remaining electrodes, and prior to and following the increments/decrements on the target electrodes. The spacing between successive incremented or decremented electrodes (4 electrodes separation) was chosen to provide a clear shift in pitch for two isolated electrodes in succession, based on pilot work with 1 implant listener.

The task used was identification of the direction of pitch change (upward or downward) resulting from the two successive increments (or decrements) applied to two electrodes. Assuming that the two pitches elicited clearly differed, this provided a measure of detection of the change applied to two of the electrodes, and hence of the concurrent segregation of the stimulation on these electrodes from that on the other electrodes (the 'background complex').

### 7.3.1 Stimuli

Total duration of all stimuli was 3.5 seconds. Increments and decrements were 1000 ms duration. The first and second increments (or decrements) started 500 ms and 2000 ms after the onset of the 5 electrode complex respectively (i.e. there was 500 ms between the end of the first and the start of the second). Schematic representations of the increments and decrements are illustrated in figure 7.1. Both the 30-ms onset *increments* and 970-ms onset *decrements* created an abrupt rise in stimulus amplitude. Conversely, the 970 ms onset increments and 30 ms onset decrements created a slow rise in amplitude. As the increments and decrements had a linear onset and offset and no plateau, the 970 ms-onset case was in effect a temporally-reversed version of the 30-ms onset case, containing *identical energy* when the same sized increment or decrement was added to the stimulation applied to a given target electrode (see figure 7.1 and table 7.1).

**Table 7.1.** Properties of increments and decrements.

| | **Onset, ms** | **Offset, ms** | **Energy rise** | **Total duration, ms** |
|---|---|---|---|---|
| 30 ms increment | 30 | 970 | Abrupt | 1000 |
| 970 ms increment | 970 | 30 | Slow | 1000 |
| 30 ms decrement | 30 | 970 | Slow | 1000 |
| 970 ms decrement | 970 | 30 | Abrupt | 1000 |

**Table 7.2.** Example of calculation of increment and decrement sizes. e = electrode number; T = threshold level; C = maximum comfortable level; DR = dynamic range; 50% point = level used for 'steady state'; 50% DR, 25% DR, 12.5% DR = increment/decrement sizes for each percentage of dynamic range. Actual levels were rounded up as required to the nearest whole number of stimulus units. N.B.: actual dynamic ranges varied between listeners (minimum = 25; maximum = 65).

| **e** | **T** | **C** | **DR** | **50% point** | **50%DR** | **25%DR** | **12.5%DR** |
|---|---|---|---|---|---|---|---|
| 3 | 100 | 155 | 55 | 128 | 28 | 14 | 7 |
| 7 | 110 | 165 | 55 | 138 | 28 | 14 | 7 |
| 11 | 120 | 172 | 52 | 146 | 26 | 13 | 7 |
| 15 | 120 | 173 | 53 | 147 | 27 | 13 | 7 |
| 19 | 115 | 175 | 60 | 145 | 30 | 15 | 8 |

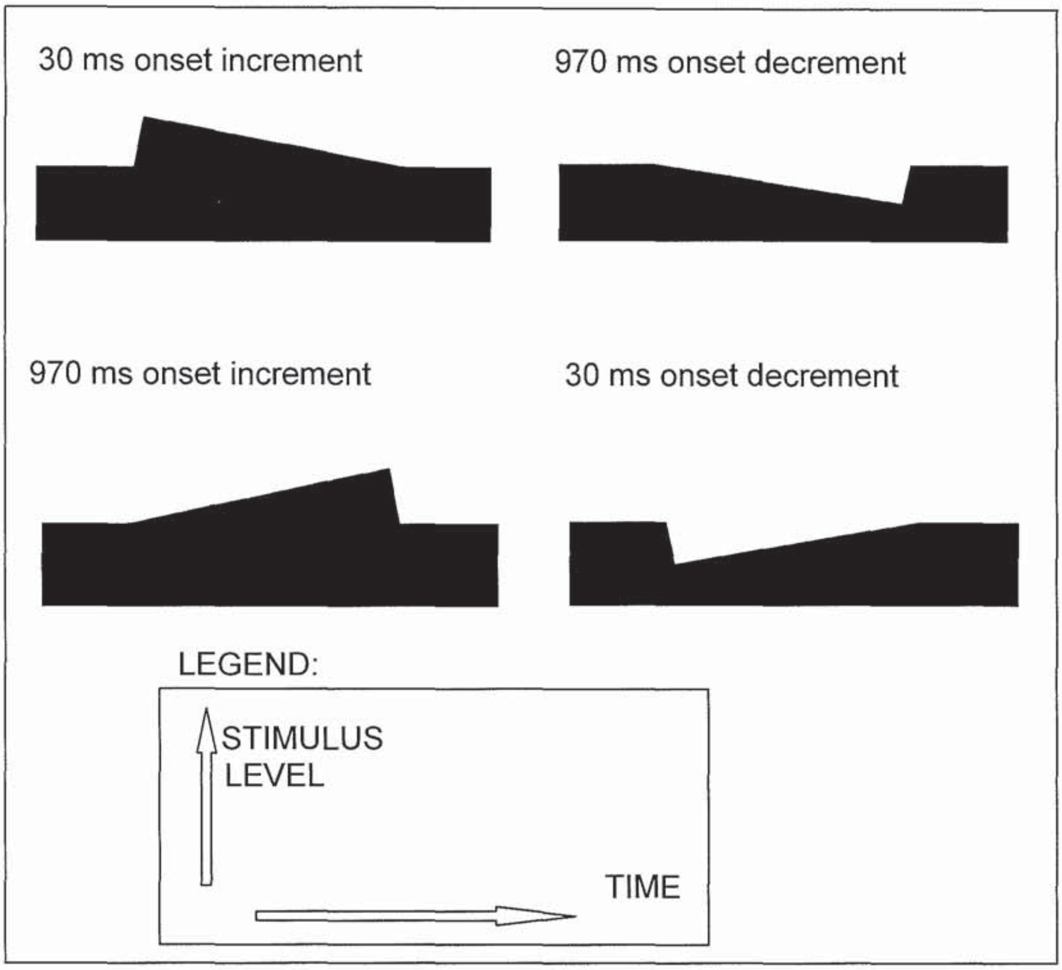
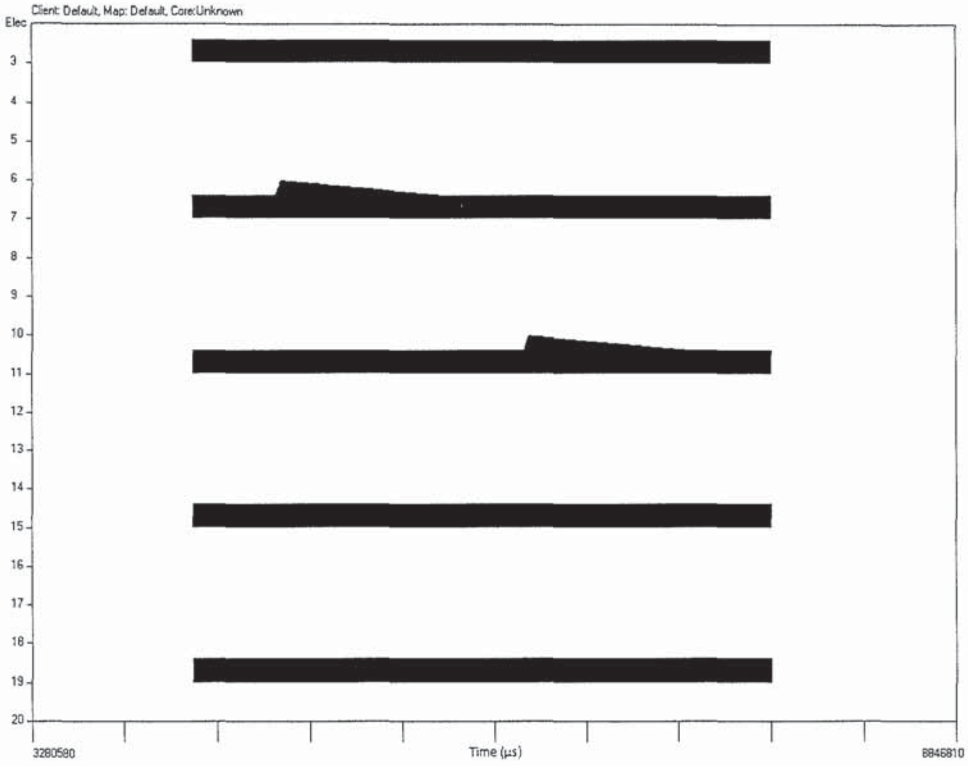


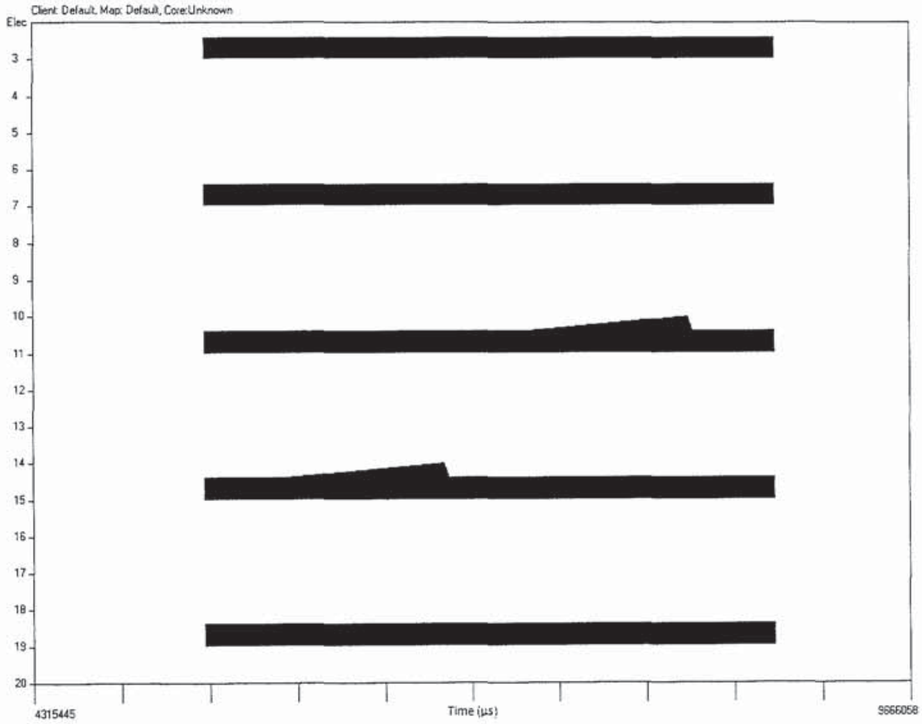
**Figure 7.1.** Schematic representations of increments and decrements applied to individual electrodes (not drawn to scale). N.B: the two stimuli shown in the upper part both have *abrupt* rises in amplitude; the two in the lower part have a *slow* rise.

In order to verify that the selected channels were activated as intended by the experimental software, all experimental stimuli were routed through the experimental speech processor and the output was analysed using a 'dummy' cochlear implant within the manufacturer's computer interface; this allowed the generation of 'Electrodograms' which illustrate visually the output on each electrode over a selected time window. Examples of these outputs are shown in figure 7.2.

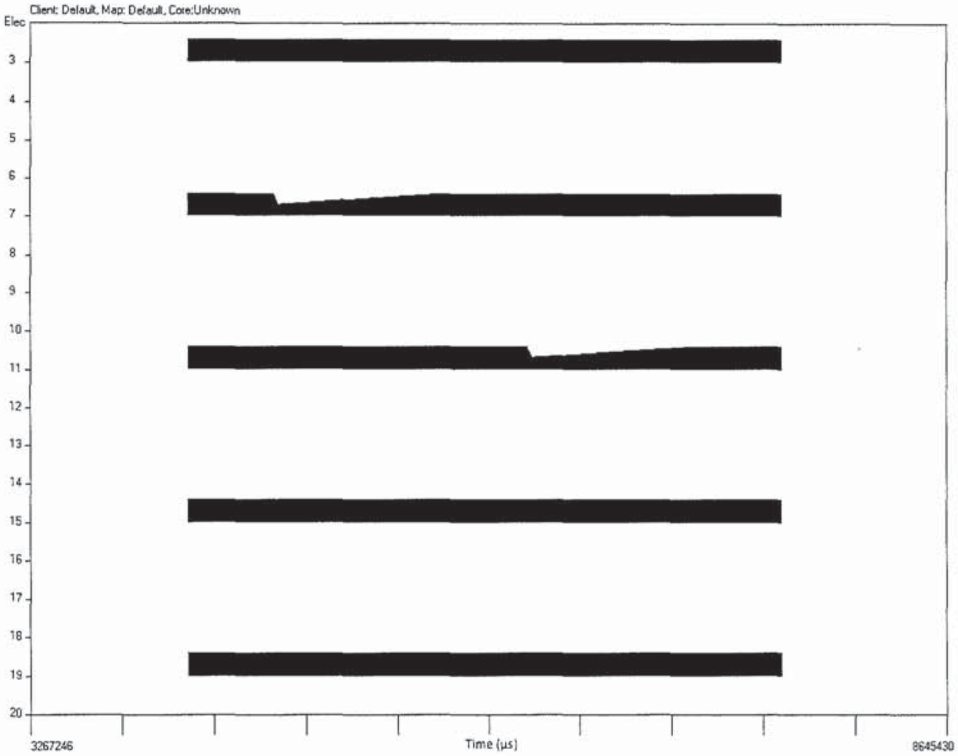
A


Client: Default, Map: Default, Core:Unknown
Elec
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
3280580
Time (µs)
8846810

B


Client: Default, Map: Default, Core:Unknown
Elec
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
4315445
Time (µs)
9666058

C


Client: Default, Map: Default, Core:Unknown
Elec
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
3267246
Time (µs)
8645430

D

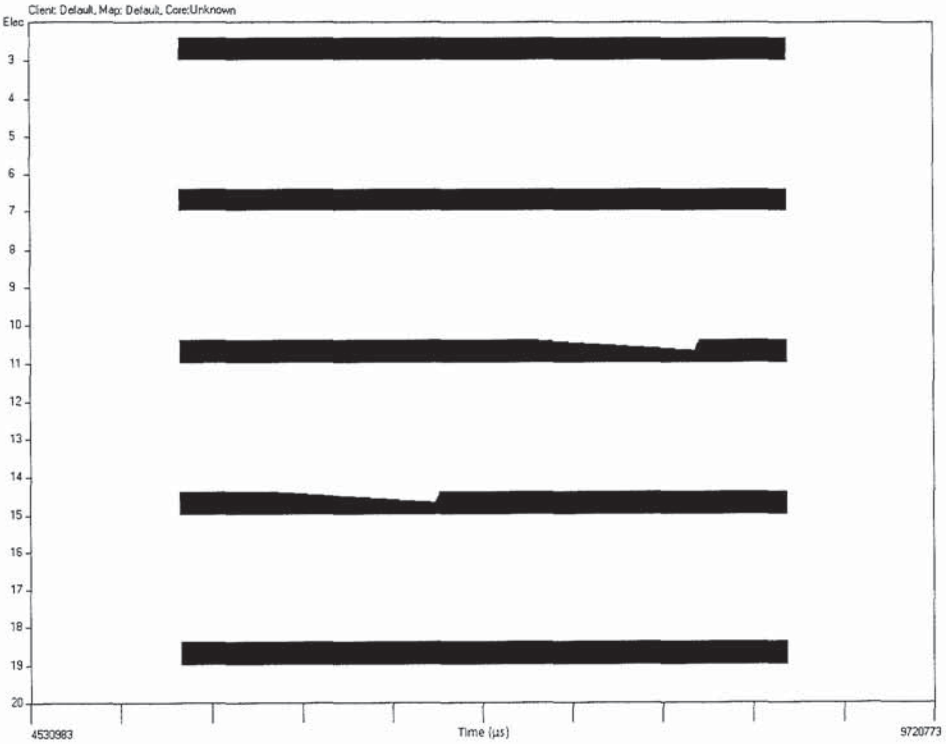

Client: Default, Map: Default, Core:Unknown
Elec
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
4530983
Time (µs)
9720773

**Figure 7.2**. 'Electrodograms' showing examples of different experimental stimuli. Filled bars illustrate when electrical stimulation was present on each electrode. The vertical height of each black bar indicates the amplitude of the stimulation applied to that electrode. The abscissa shows time elapsed in µs. The ordinate shows electrode number; lower numbered electrodes towards the top (i.e. more basal/higher pitched) and higher numbers towards the bottom (more apical/lower pitched). The pitch shifts illustrated in these examples are therefore downwards in panels A and C, and upwards panels B and D.

(A): Increments with 30 ms onsets; e7 followed by e11.

(B): Increments with 970 ms onsets; e15 followed by e 11.

(C): Decrements with 30 ms onsets; e7 followed by e11.

(D): Decrements with 970 ms onsets; e15 followed by e11.

### 7.3.2 Listeners

Listeners were 8 experienced adult users of the Nucleus CI24 device. All had full electrode insertions and showed no evidence of electrode pitch reversals. Demographic and other details are shown in Table 7.3.

Table 7.3. Demographic and clinical details of listeners. Those marked with an asterisk had not taken part in any previous experiments.

| Listener | Age | Gender | Open set speech recognition score (BKB sentences in quiet, % correct) | Pulse rate (pulses per second) in normal daily use | Pulse width (µs) in normal daily use |
|---|---|---|---|---|---|
| L4 | 33 | F | 88 | 250 | 25 |
| L12* | 69 | F | 93 | 1200 | 25 |
| L13 | 51 | F | 100 | 900 | 25 |
| L15* | 69 | M | 68 | 900 | 25 |
| L16* | 64 | M | 53 | 500 | 25 |
| L17* | 79 | F | 93 | 1200 | 25 |
| L18* | 73 | M | 27 | 900 | 25 |
| L19* | 63 | M | 74 | 2400 | 25 |

### 7.3.3 Procedure

A single-interval 2AFC design was used. Increments and decrements were tested separately, in parts 1 and 2 of the experiment respectively. Thus stimuli were blocked first by increment/decrement then by size. Increments were tested first because it was found during pilot testing that listeners found the stimuli much easier to detect. Larger increment sizes (in part 1) and decrement sizes (in part 2) were tested first as listeners were able to detect them more easily. In each case, testing only progressed to the next smaller increment (or decrement in part 2) when listeners demonstrated above-chance performance for the current stimulus size. In each experimental block, the sequence of electrodes on which the increments/decrements were applied was randomised between 4 possible orders: LM, ML, MH, or HM (e7=H, e11=M, e15=L; equal numbers of each order per block). Each experimental block consisted of 10 repetitions of each possible order, for each combination of onset time (30 or 970 ms) and increment (or decrement, in part 2) size (50%, 25%, or 12.5% of DR). All listeners were tested initially with training stimuli in which the steady-state level for all 5 electrodes was set at T–level (threshold), and increments of 100% of dynamic range were applied. This provided familiarity with the task, using stimuli that were easily distinguishable. Listeners only proceeded to the experiment itself when they were able to reliably achieve 100% correct performance in this condition after training.

### 7.3.4 Task and responses

The purpose of the experiment was explained to the listeners; they were instructed both verbally and in writing. The task was to attend to each 3.5-second complex sound and listen for two brief pitches, one higher than the other, occurring during the longer sound. They then had to report the order of the pitches, i.e. did they hear the pitch pattern go up or down (similar to the task instructions used by Bregman et al.,1994). Each listener was seated in front of a response box on which one button marked S was illuminated; they were instructed to press the lit button to hear the next sound, which was presented 500 ms after the button was pressed. After each sound, they responded by pressing one of two buttons numbered 1 or 2, to indicate whether they had heard a falling or rising pitch pattern. Having responded, the S button lit up again ready for the next sound. This continued until the experimental run was complete. All listeners received training and practice on the task prior to experimental runs. The experimental software recorded each response and whether it was correct or not, but no feedback was provided on the accuracy of responses.

## 7.4 Results

### 7.4.1 Mean results: Increments

Mean results for all 8 listeners are shown in figure 7.3. For the 50% DR condition, mean scores were significantly above chance for both 30-ms and 970-ms onset increments (95% confidence intervals below the means were 66.1 and 63.6 respectively). The mean score for the rapid rise time (30 ms) was slightly higher (77.8%) than for the slow rise time (970 ms) (73.9%), but this difference was not significant (t(7)=0.81; p = 0.44). Means in the 25% DR and 12.5% DR conditions, for both onset times, were all above chance (but not significantly; 95% confidence intervals below the means were all below 50%). Again, there was no significant difference between the rapid (30 ms) and slow (970 ms) onsets for either of the two smaller increment sizes. Overall, the mean data suggest that listeners were generally only able to perform above chance with increments of 50 % of their dynamic range; with smaller increments, mean performance fell towards chance levels.

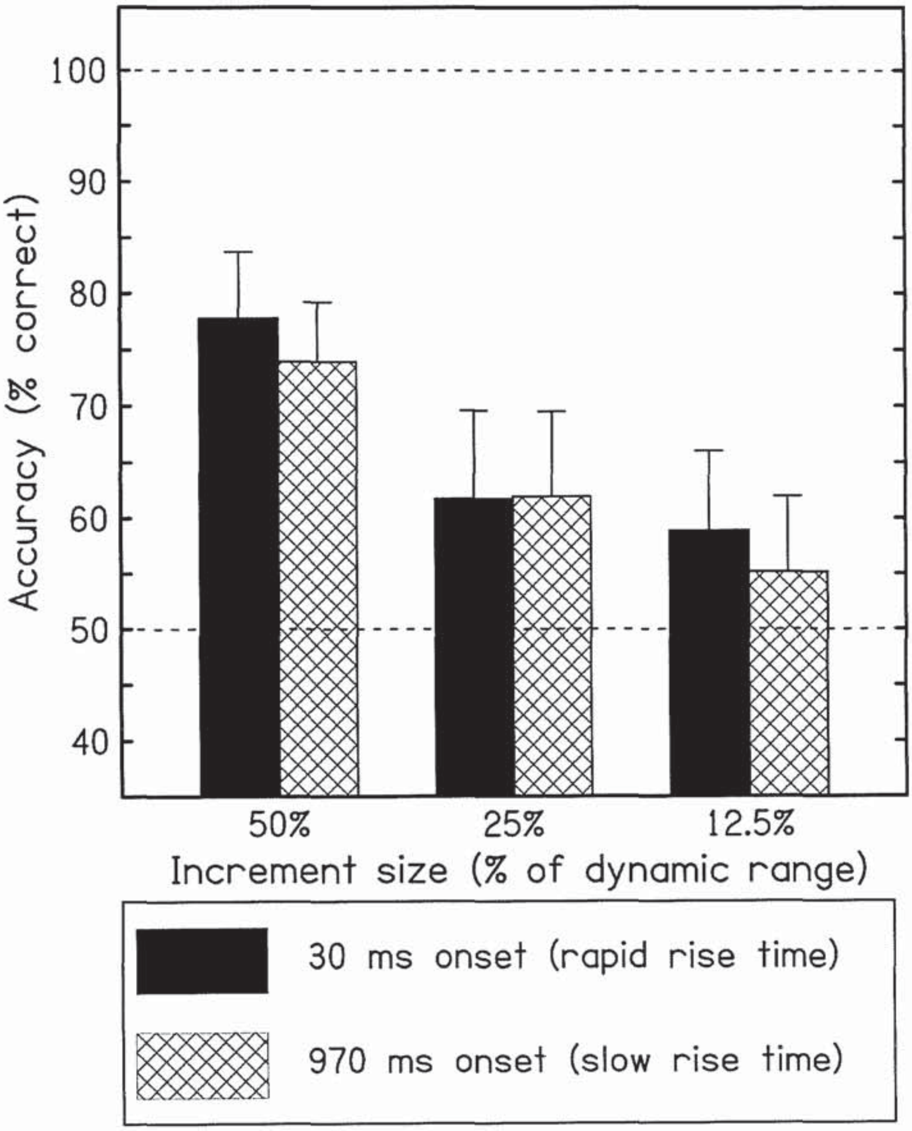


**Figure 7.3**: Mean data for all 8 listeners: Increments.

The lower dashed line indicates chance level. N.B: for the purposes of illustration, a value of 50% was inserted into the calculation of means where data were absent and chance-level performance can be assumed. Inter-subject standard errors are indicated by vertical lines above each mean.

A two-way within-subjects ANOVA on the increments data showed a significant main effect of increment size on accuracy of pitch-change judgements (larger increments were associated with greater accuracy): F(2,10)=7.85, p=0.009. There was no significant main effect of onset time: F(1,5)=2.94, p=0.147, although a trend was apparent for shorter onset times (i.e. more rapid rise times) to be associated with greater accuracy. There was no significant interaction between onset time and increment size: F(2,10)=0.396, p=0.683. [N.B: listeners L18 an L19 were unable to complete the task in the 12.5% of dynamic range increment size condition; their data for that condition were not included in the analysis.]

The ANOVA was repeated using arcsine-transformed data, with the intention of reducing the heterogeneity of the data. The results, which confirmed the outcome of the first analysis, were as follows:

Main effect of increment size: F(2,14)=7.584; p=0.006; *significant*

Onset time: F(1,7)=1.779; p=0.224; N.S.

Increment size x onset time: F(2,14)=1.628; p=0.231; N.S.

As the results for one listener (L13) were quite different from the others, the ANOVA was also repeated excluding the data for L13. The results, which are consistent with the outcome of the first analysis, were as follows:

Main effect of increment size: F(2,8)= 10.885, p=0.005; significant

Onset time: F(1,4)=3.078, p=0.154; N.S.

Increment size x onset time: F(2,8)=0.353, p=0.713; N.S.

### 7.4.2 Individual results: Increments

Considerable variability in the results for individual listeners was apparent, as shown in figure 7.4. Only 4 out of the 8 listeners showed performance above chance for the largest increment size (+50%DR), and performance generally deteriorated for the smaller increments. Two listeners (L12 and L16) showed an apparent advantage for a rapid rise-time for all three increment sizes, although this difference was not significant (t test). Two listeners (L15, L17) showed some advantage for more rapid rise times for some increment sizes but not others. Their performance was at or near chance for the two smaller increment sizes, and only above chance for the largest increment (+50%DR). Two others (L18 and L19) showed a reversal of the expected pattern (with regard to rise time), with better performance for 970 ms onset increments, and were unable to perform the task at all for the smallest increment size (+12.5% of dynamic range). Results for one listener (L13) differed markedly from the others; she showed near-perfect performance for both onset times at all increment sizes; this near-ceiling performance may have obscured any effect of rise time. Results of further testing of this exceptional listener are described below.

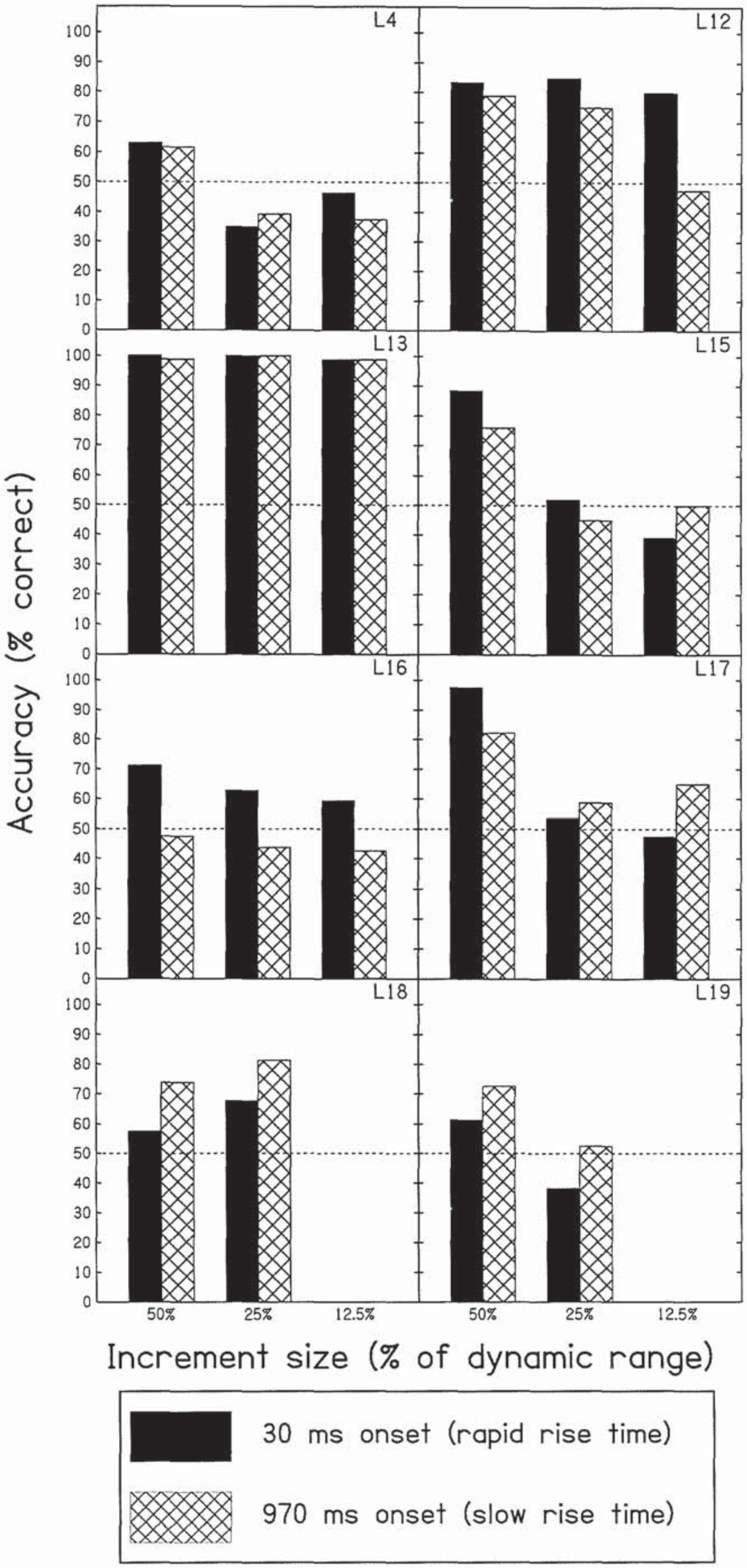

L4
L12
L13
L15
L16
L17
L18
L19
Accuracy (% correct)
0
10
20
30
40
50
60
70
80
90
100
50%
25%
12.5%
Increment size (% of dynamic range)
30 ms onset (rapid rise time)
970 ms onset (slow rise time)

**Figure 7.4.** Individual results for all 8 listeners: Increments. Dashed lines indicate chance level. Absent columns are where listeners were unable to hear any pitch changes and had responded at chance level, or were unable to continue with the task.

### 7.4.3 Mean results: Decrements

Mean results for the decrement conditions are shown in figure 7.5. Mean performance for the slow rise time condition (30 ms onset) was very close to chance for all three decrement sizes.

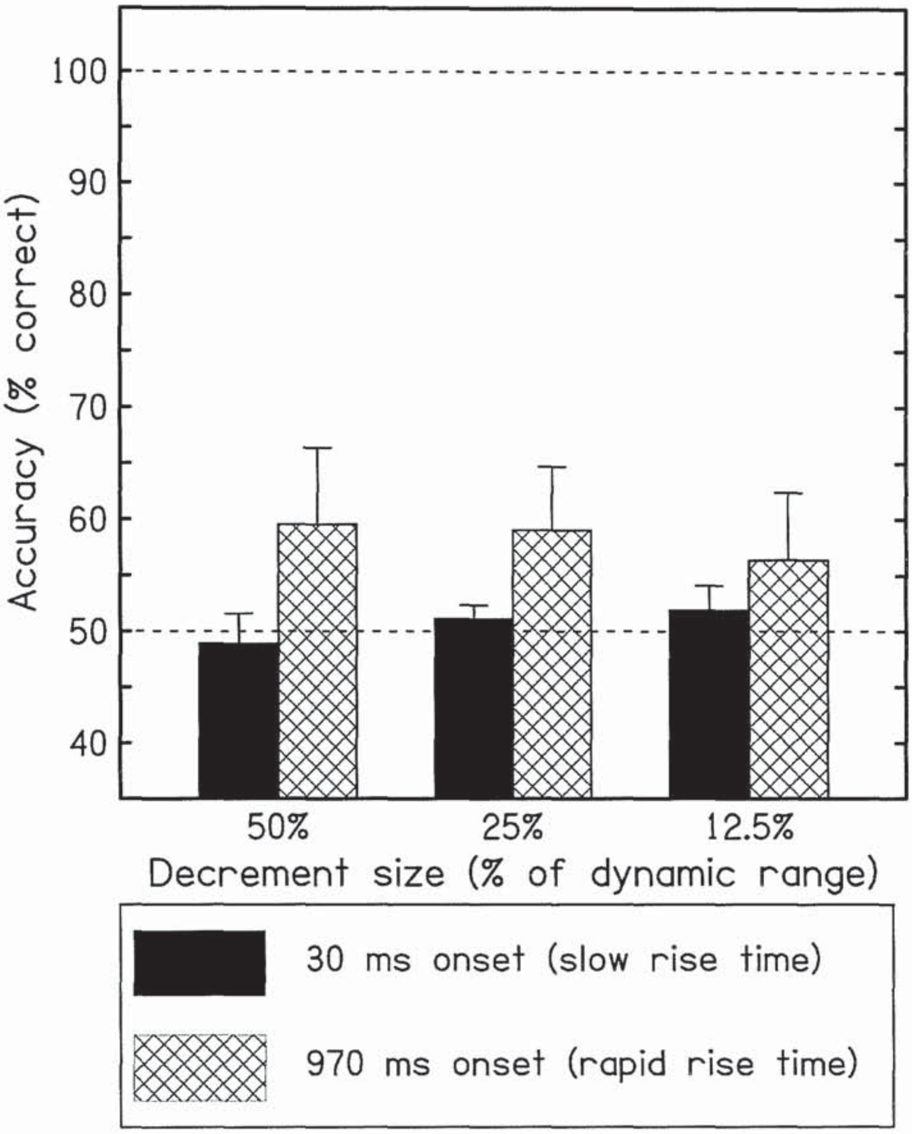


**Figure 7.5**: Mean data for all 8 listeners: Decrements. Dashed lines indicate chance level. N.B: for the purposes of illustration, a value of 50% was inserted

into the calculation of means where data were absent and chance-level performance can be assumed. Inter-subject standard errors are indicated by vertical lines above each mean.

For the rapid rise times (970 ms onset), mean performance for all three decrements sizes was slightly above chance (but not significantly; 95% confidence intervals below means were 46.3, 48.9 and 30.2 respectively). In the 50% DR condition, data were available for both rise times for 7 listeners; the mean score was higher for the rapid rise time (970 ms onset) than for the slow rise time (30 ms onset) (59.6% vs. 48.8%), but this difference was not significant (p=0.085). In the 25% DR condition, data were available for both rise times for only 4 listeners; again, although the mean score was higher for the rapid rise time (970 ms onset), the difference was not significant (p=0.275). For the smallest decrement size (12.5%DR) data were only available for both onset times for two listeners and so statistical analysis was not attempted. In view of the proportion of missing data across conditions, analysis of variance was not attempted due to concern that it would provide invalid results.

### 7.4.4. Individual results: Decrements

Individual results for the decrements are shown in figure 7.6. Six out of the 8 listeners were unable to perform above chance in any condition completed; 4 out of 8 were unable to perform the task at all for the 25%DR decrements (i.e. they responded entirely randomly and further testing was discontinued), and 6 out of 8 were unable to perform it at the smallest decrement size (12.5%DR). The results for L13 again differed markedly from those for the others; for all three decrement sizes she showed near-perfect performance for 970-ms onsets (rapid amplitude rise) and near-chance performance for 30-ms onsets (slow rise), and this difference was significant (t=0.025). L16 showed a similar advantage for the 970-ms onset, but only at the largest decrement size (50% of dynamic range). L12 and L15 appeared to show some advantage for the 970-ms onset (more convincingly for L12), although still performing at or near chance levels.

### 7.4.5 Further testing of the exceptional listener (L13)

For both increments and decrements, the results for one listener (L13) were very different from those for the other 7; she appeared to behave more like a NH listener. In order to explore her performance limits on this task further, she was tested with smaller increment and decrement sizes than previously used. The task and procedure were identical to those used before. Increments and decrements of 4, 2 and 1 stimulus units were tested. These were equivalent to approximately 6%, 3% and 1.5 % of dynamic range (dynamic ranges for this listener were uniform across the three target electrodes). It should be noted that at small increment sizes (below about 4 stimulus units), the difference between a rise time of 30 ms and 970 ms essentially becomes meaningless. At the limit, an increment (or decrement) of 1 stimulus unit was a step function with no difference between the 30 ms and 970 ms conditions. However, increments or decrements of 4 units or greater provided a reasonable approximation to the intended amplitude envelopes.

Results are shown in figure 7.7. Once again, her performance with increments was near-perfect for increments of 4 or 2 stimulus units, and only slightly reduced (i.e., still well above chance) for a 1 stimulus unit increment. As before, no clear effect of onset time was revealed due to the near-ceiling performance. For the decrements task, her performance for a 30-ms onset (slow rise) was again close to chance for all 3 decrement sizes. However, for 970 ms-onset decrements (rapid rise), a clear effect of decrement size was apparent; she produced scores of 100% and 81% correct for decrements of 4 and 2 stimulus units, respectively, falling to below chance (47% correct) only for the smallest decrement (1 stimulus unit). Thus, the results for L13 (figure 7.7) were closest to what might have been expected for NH listeners, although any effect of onset time for increments was obscured by her remarkably good performance - even for increments as small as 1 stimulus unit. She was apparently unable to detect decrements with rapid (30 ms) onsets, i.e. with slow rises in level, but she was able to reliably detect decrements with rapid level increases (970 ms onsets) almost as well as rapid–onset increments.

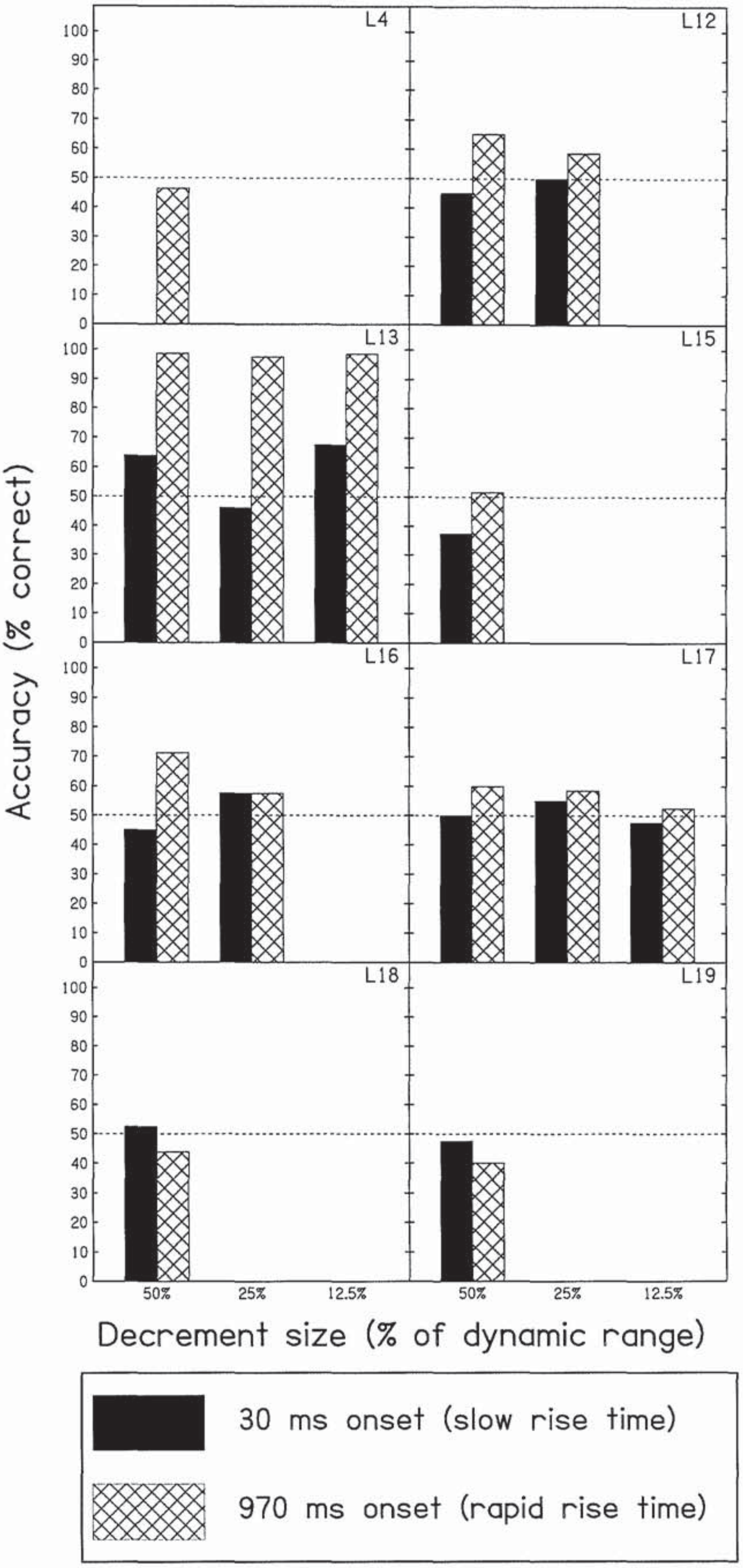

L4
L12
L13
L15
L16
L17
L18
L19
100
90
80
70
60
50
40
30
20
10
0
Accuracy (% correct)
50%
25%
12.5%
Decrement size (% of dynamic range)
30 ms onset (slow rise time)
970 ms onset (rapid rise time)

**Figure 7.6.** Individual results for all 8 listeners: Decrements. Dashed lines indicate chance level. Absent columns are where listeners were unable to hear any pitch changes and had responded at chance level, or were unable to continue with the task.

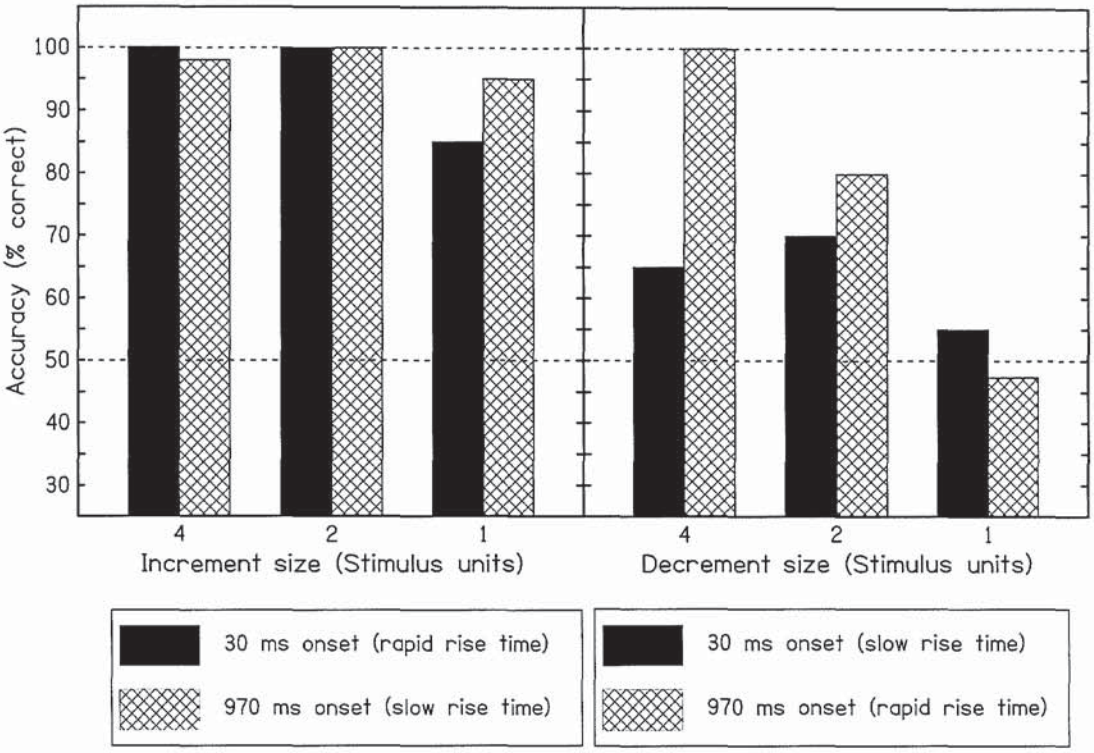


**Figure 7.7**: Further results for listener L13. The lower dashed lines indicate chance level. The left panel shows results for increments, and the right panel for decrements. In both cases black columns indicate 30-ms onsets and cross-hashed columns indicate 970-ms onsets.

## 7.5. Discussion

The aim of this experiment was to evaluate the ability of cochlear implant listeners to use a) temporary increases in energy or b) abrupt increases in level (with or without an increase in energy) on individual electrodes as a cue for the perceptual segregation of one component from a complex (comprising simultaneous stimulation on multiple electrodes). Kubovy and Daniel (1983) revealed that 'momentary amplitude disparity' can be used for concurrent pitch segregation. They showed that a momentary reduction in amplitude (for 80 ms) in one tone within a complex, followed a return to its former intensity, can cause the pitch of that tone to 'pop out', even though there has been no actual increase in energy. Bregman et al. (1994b) showed that abrupt *increases* in energy on single tones within a complex could have the same effect.

The results suggest that some implant listeners are able to detect temporary increases in stimulus level to pick out components from a complex tone, but they do not appear consistently to show the same benefit of suddenness of onset as seen in normal-hearing listeners. A possible explanation is that in normal-hearing listeners, release from adaptation of inner hair cell responses may contribute to the 'pop out' effect associated with sudden changes. This would not be apparent in cochlear implant listeners.

Also, they generally require much larger increases in stimulus amplitude, compared to normal-hearing listeners who are able to detect amplitude changes as small as 1 dB in a similar task (Bregman et al., 1994b). One listener (L13) showed performance that differed from that of the others; she behaved more like a normal-hearing listener. She showed near-perfect detection of increments as small as 1 stimulus unit. It is difficult to be sure why her performance differed so markedly from that of the other listeners on this task. The fact that she displayed the best open-set speech recognition with her implant (100% correct) out of all 8 listeners tested is further evidence of her exceptional discrimination ability. She also showed a clear advantage for sudden *offsets* in decrements (i.e. rapid amplitude rises) on individual electrodes. Thus, she was apparently able to use *either* small, temporary increases in energy *or*, in the absence of an energy peak, abrupt rises in level, as cues for segregation.

The generally poor performance across listeners for decrements in stimulus amplitude, combined with the absence of any clear overall effect of increment envelope, suggest that most cochlear implant (CI) listeners (aside from exceptional individuals) rely mainly on detecting peaks in energy and are unable to use abruptness of change as a cue for concurrent sound segregation, unlike NH listeners (Kubovy and Daniel, 1983; Bregman et al., 1994b). The study by Carlyon et al. (2007) suggested that onset asynchrony may be available to CI listeners as a cue for concurrent segregation; however, the effect they reported was small and they found no evidence that differences in pulse-rate or pulse asynchrony could be used for simultaneous grouping. Thus, evidence so far suggests that cochlear implant listeners have access to few, if any, of the cues for concurrent sound segregation that NH listeners take advantage of in difficult listening situations with mixtures of competing sounds.

# Chapter 8: Experiment 6

## 8.1 Introduction

All currently available cochlear implants use intra-cochlear arrays with multiple electrodes, intended to take advantage of the tonotopic arrangement of the cochlea, and endeavour to convey spectral information about speech and other sounds through selective stimulation of discrete sections of the neural population located in the modiolus. A perfectly designed and functioning implant system would therefore have electrodes that stimulated focused, discrete points within the cochlea, providing a high degree of frequency selectivity and distinct pitch percepts from one electrode to the next. Better electrode discrimination has been shown to correlate with improved speech recognition; for example, Donaldson and Nelson (2000) reported a positive correlation between place-pitch sensitivity and consonant place-cue performance in users of the Nucleus CI-22 implant. However, in reality, activation of electrodes spatially close to each other leads to stimulation of overlapping neural populations and therefore to channel interactions. This inevitably leads to a reduction in the quality of spectral information that is conveyed, by creating spectral smearing and thus limiting the performance of the implant. Some evidence for this was reported by Litvak et al. (2007a). They measured the spectral modulation threshold or SMT (defined as the smallest detectable spectral contrast in a spectral ripple stimulus) for implant listeners and also for normally hearing listeners with vocoder simulations. They attempted to simulate varying degrees of current spread by adjusting the slope of drop-off of filters for filtered noise bands. They found that vowel and consonant identification scores for the vocoder listeners dropped significantly with reduced slope in the noise-band filters, accompanied by increased SMT thresholds. Also, the reductions in speech recognition thus created in their vocoder listeners were similar to those found in their implant listeners with similar SMTs; based on this, they concluded that variability in the spread of neural activation largely accounts for the variability in speech recognition in implant listeners.

Various attempts have been made to alleviate this problem with varying degrees of success, for example: interleaved stimuli as in the CIS (Continuous Interleaved Sampling) strategy (e.g. Wilson, 2006); bipolar (instead of monopolar) stimulation where the current path from active to return electrodes is shortened; 'tripolar' stimulation in which current of opposite polarity is simultaneously delivered to two electrodes adjacent to the active electrode (e.g. Litvak et al., 2007b); and electrode arrays that position electrodes closer to the modiolus (e.g. the Nucleus 'Contour' or Advanced Bionics 'Helix' arrays). Nonetheless, spread of excitation in the cochlea, with resulting loss of frequency selectivity, remains a significant issue for implant design and performance.

Reduced spectral information is thought to account for many of the known limitations of performance with implants, and is likely to be a contributing factor to the generally poor performance in noise, as well as to the apparently poor ability of implant listeners to perform well on tasks requiring stream segregation, as evidenced by the results of the previous chapters. It is therefore important to find reliable and meaningful ways to evaluate channel interaction and spread of excitation, and if possible to investigate links between these and other measures of implant performance. Such information might be used to understand how speech processing strategies could be developed to improve performance. Psychophysical approaches such as forward masking (e.g. Chatterjee and Shannon, 1998) or gap detection (e.g. Hanekom and Shannon, 1998) have been investigated, as discussed in Chapter 2. However, little evidence of a relationship between such measures and speech recognition ability has been reported to date, and as pointed out by Cohen et al. (2003), it is unclear how much psychophysical measures of channel interaction might be influenced by central processing effects.

An alternative, and potentially a more reliable and quicker approach, is to use the Neural Response Telemetry (NRT $^{TM}$) system for recording the electrically Evoked Compound Action Potential (ECAP) in the Nucleus CI24 cochlear implant. NRT recording involves presentation of a stimulus to one electrode in the intra-cochlear array and recording of the ECAP response generated as a result from another, usually neighbouring electrode. Application of stimulus pulses to a 'probe' electrode creates a current field in the area surrounding it. This creates an excitation field, within which neurons fire and generate an action potential. The number of neurons firing decreases to zero with increasing distance from the stimulating electrode. The recording electrode records the sum of the action potentials from the population of neurons surrounding it, which creates the ECAP response. Responses to a sequence of repeated stimuli are averaged by the recording software, which reduces recording noise to an acceptable level (there remains a 'noise floor' which obscures very small responses). The subtraction method used in the NRT software removes the large stimulus artefact, allowing recording of the neural response alone. The basis for the recording of the neural response using the subtraction method is shown in figure 8.1. In effect, four separate recordings are made of the response from a specific electrode, which are then processed and averaged by the software to produce a measurable neural response. When the probe stimulus occurs without the preceding masker pulse, the result is a response with both stimulus artefact and the neural response present (A). When the probe stimulus is preceded by a masker pulse, usually on the same electrode, and provided the delay between masker and probe is sufficiently short (e.g. 500 µs), the effect is to cause most or all of the nerve fibres in the vicinity of the probe electrode to be in a refractory state, so that they do not respond to the probe pulse that follows (B). The neural response on its own is then obtained in the recording software by subtraction, according to the formula A-(B-(C-D)), where C is the response resulting from the masker pulse on its own and D is the result of a 'switch-on artefact', i.e. a condition where the implant is 'powered up' but no stimulation occurs. Subtraction of the 'B' response (which contains both a probe stimulus artefact and masker stimulus artefact) from the 'A' response results in the substantial reduction of the probe stimulus artefact. Further subtraction of the 'C' response results in the substantial reduction of the masker stimulus artefact, and subtraction of the 'D' response substantially reduces the 'switch-on' artefact. The resulting ECAP, which is obtained by averaging over a number of A/B/C/D stimuli, consists of a negative peak (N1) followed by a positive peak (P2) (see figure

8.2). The amplitude of the response can be measured by calculating the peak-to-peak amplitude difference between N1 and P2, in μV.

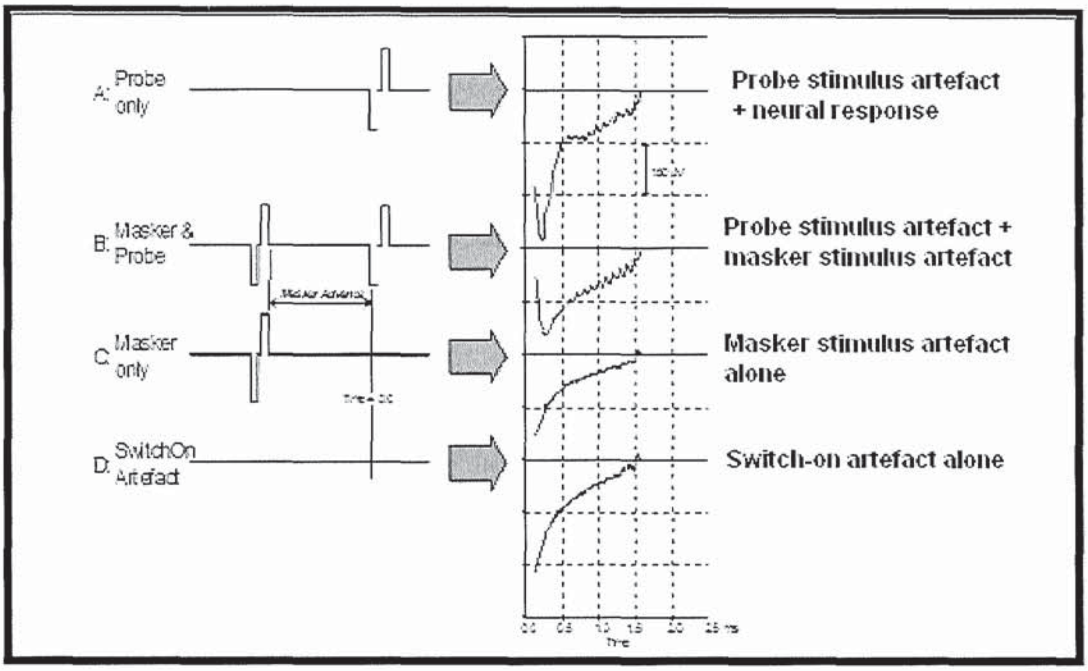
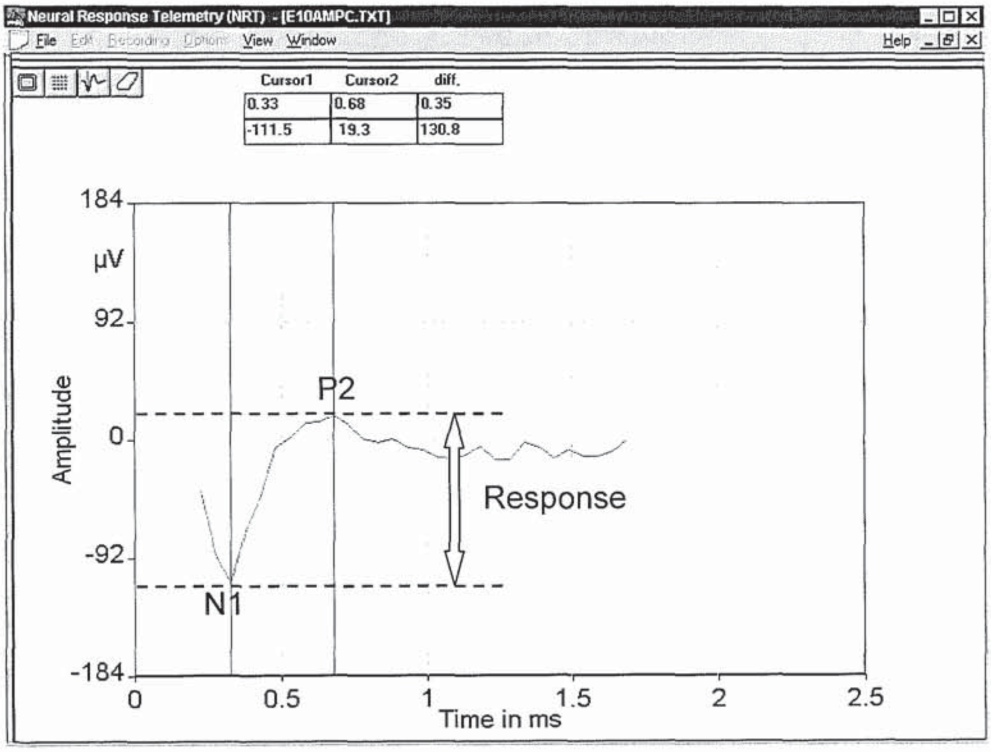


**Figure 8.1**. Schematic representation of the subtraction method for recording the ECAP response

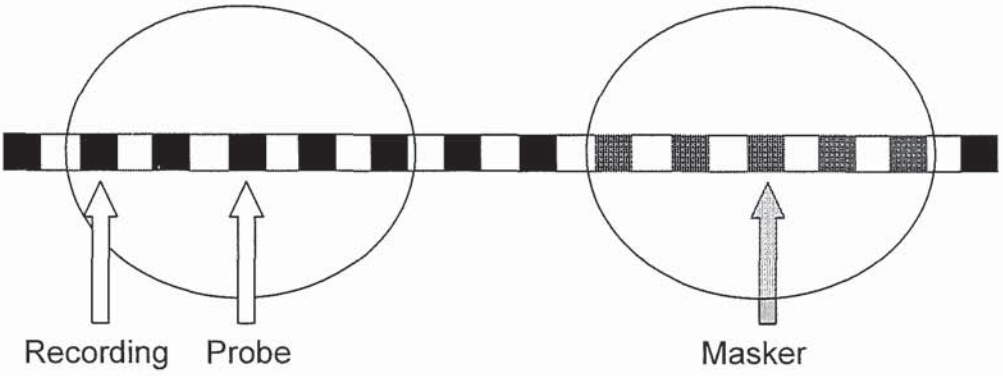


| Cursor1 | Cursor2 | diff. |
|---|---|---|
| 0.33 | 0.68 | 0.35 |
| -111.5 | 19.3 | 130.8 |

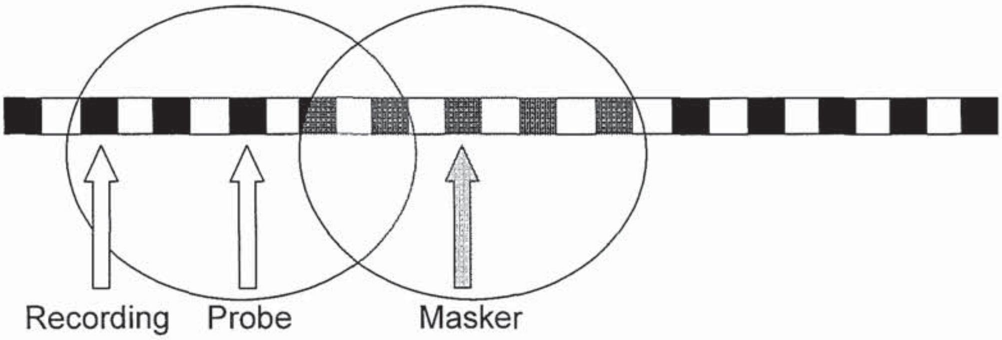
**Figure 8.2**. Example of a typical ECAP response showing measurement of response amplitude, as displayed in the Nucleus NRT 3.1 software. The abscissa shows time in ms, measured from the stimulus pulse. The ordinate shows the instantaneous amplitude of the ECAP response in µV. The overall size of the response is estimated from the peak-to-peak amplitude between the first negative peak (N1) and the following positive peak (P2). In this example, the peak-to-peak amplitude of the ECAP response is 130.8 µV, as shown in the panel near the top. Vertical lines at 0.33 and 0.68 ms indicate the latencies of N1 and P2.

In the subtraction method, the largest ECAP response can be expected when the masker and probe pulses stimulate the same electrode, as presumably identical populations of local neurons are excited and so the clearest and largest neural response is recorded. Conversely, if the masker stimulus is applied to an electrode spatially distant to the probe electrode, little or no neural response will be recorded as there is no overlap between the neural populations excited by these stimuli. Thus, the size of the ECAP recorded using this method can provide an estimate of the degree of overlap between the neural populations stimulated by masker and probe. By maintaining the probe stimulus on one electrode and moving the masker between other electrodes in the array on each recording, a measure of spread of excitation can be obtained; Cohen (2003) referred to this as the Variable Masker, Fixed Recording location ECAP spatial measure, or VMFR. This is illustrated in figure 8.3.

(A) No overlap between probe and masker excitation patterns: no ECAP recorded



(B) Partial overlap between probe and masker excitation patterns: reduced amplitude ECAP recorded



(C) Complete overlap between probe and masker excitation patterns: maximum amplitude ECAP recorded.

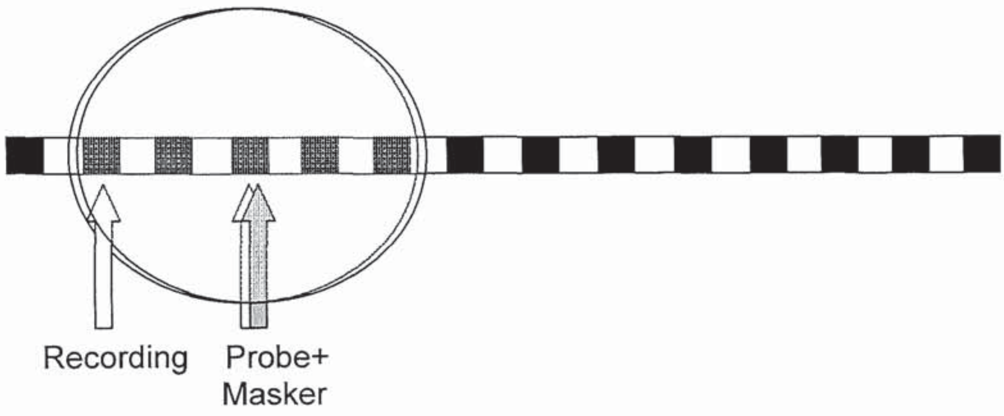


**Figure 8.3**. Schematic representation of the fixed probe, variable masker position method for recording spread of excitation. A section of the electrode array is shown. Individual electrodes are illustrated by black rectangles. Areas of surrounding neural tissue stimulated are shown by the grey ovals. In the

first example (A), probe and masker electrodes are spatially distant and there is no overlap of stimulation; no ECAP response is recorded. In example (B), probe and masker are closer together and there is partial overlap of the areas of excitation produced by each: an ECAP response may be recorded, but its amplitude will be below maximum. In example (C), probe and masker are on the same electrode; the maximum amplitude ECAP is recorded as identical areas of neural tissue are stimulated.

Thus, the VMFR method should provide an objective measure of the spread of excitation in the cochlea resulting from stimulation of a single electrode. Narrow, selective stimulation of a discrete area in the cochlea in the immediate vicinity of the probe electrode should lead to a rapid reduction in recorded ECAP amplitude as the masker electrode is moved along the array in either direction away from the probe electrode. In contrast, stimulation that causes excitation of a wide area of neural tissue beyond the probe electrode should lead to a less rapid reduction in ECAP size as the distance between masker and probe electrodes is increased. Based on this, the width of the VMFR profile, obtained by plotting the ECAP amplitude against masker electrode location, should provide a direct measure of the spread of excitation in the cochlea: wider profiles indicate greater spread of excitation and vice-versa. Good agreement between profiles obtained in this way and those obtained from psychophysical forward masking were reported by Cohen et al. (2003).

The first aim of this experiment was to evaluate the range of ECAP-VMFR profiles that are typically seen in a group of implant users with more than one type of electrode array (e.g. either straight arrays or the pre-curved 'Contour' array), and who have a broad range of speech recognition performance (including both good and poor performers). Secondly, any possible correlations with results of other experiments described in this thesis are explored. Reduced spread of excitation should be associated with reduced channel interaction, and so with more distinct pitch percepts; thus, smaller ECAP profile widths might be expected to be associated with better performance on tasks that rely on perceptual differences between electrodes. If this is the case, then significant correlations should be apparent between ECAP spread of excitation profiles and profiles of performance obtained on perceptual tasks where electrode separation is an independent variable.

## 8.2 Method

### 8.2.1 Overview

ECAP responses were recorded using the Nucleus NRT (Neural Response Telemetry $^{TM}$) system and NRT software (version 3.1). Recordings were made using an experimental 'Sprint' speech processor compatible with the NRT software and a standard clinical programming setup. This allows telemetry between the implanted intracochlear device and the recording software via the speech processor and computer interface. Listeners' own speech processors were not used or altered in any way.

### 8.2.2 Stimuli

Stimuli consisted of biphasic current pulses at a fixed rate of 35 Hz; pilot work had demonstrated that this pulse rate produced robust NRT responses. The masker-probe interval (i.e. the time interval between the masker and probe pulses) was maintained at 500 μs throughout; pilot work and clinical experience had demonstrated that this value provided the most robust ECAP responses. A monopolar stimulation mode (MP1+2) was used, i.e. the active stimulated electrode was paired with remote reference electrodes located on the implant package and the separate return electrode. Monopolar stimulation is most commonly employed in clinical use; also, previously published NRT spread of excitation data have been based on monopolar stimulation. A recent, small study by Busby (2007, personal communication) showed that although NRT spread of excitation can be recorded using bipolar stimulation, the recording parameters are more difficult to define and the results less clear than with monopolar stimulation. Electrode 11 was used as the probe electrode throughout. This was chosen a) because it is located in the centre of the electrode array, allowing considerable scope for movement of the masker electrode away in both basal and apical directions and b) to facilitate comparison with the results of experiments 1 and 2 reported in this thesis, for which the fixed stimulus (Tone A) always stimulated electrode 11.

Stimulus levels were at Loudest Acceptable Presentation Level (LAPL); this differs from, and is normally slightly higher than the C level (comfort level) as it represents the highest *acceptable* level as distinct from the highest *comfortable* level. The absolute level used varied from listener to listener and was measured in the condition where probe and masker stimulated the same electrode (e11). The same level, expressed in stimulus units, (within each implant user) was then used throughout the recording session. Pilot work had demonstrated that robust NRT responses were most reliably obtained at high stimulus levels, and a previous study (Cohen at el, 2003) reported that stimulus level had little effect on the widths of ECAP functions. Also, the aim was to obtain a large, clear ECAP response when masker and probe coincided on the same electrode, so that the resulting reduction in response size could be clearly observed as the response amplitude decreased with increasing probe/masker separation. Masker and probe stimuli were presented at the same stimulus level; pilot work had also shown that this combination of parameters was successful in recording ECAP responses.

### 8.2.3 Procedure

Listeners were instructed that they would hear a series of sounds, some of which would be loud but should not reach their loudness discomfort levels. No subjective responses were required, beyond the precaution of indicating if any of the sounds heard were painful or otherwise uncomfortably loud. Each series of recordings took around 30 to 45 minutes to complete. Each individual recording took around 30 seconds to one minute. NRT responses were first recorded in the conventional way, i.e. with masker and probe stimuli both applied to the same electrode (e11 in this case). Electrode 13 was in most cases used as the NRT recording electrode; pilot work and clinical experience had shown that this separation between probe and recording electrode was most likely to provide clear ECAP responses, and provided the best compromise between response size and proximity between probe and recording electrodes. LAPL was established by gradually increasing the stimulus levels until the listener indicated that the loudest acceptable level had been reached (the software allows a 'stimulus only' condition in which the probe/masker stimuli are presented but no recording occurs). This level was then used throughout subsequent recordings. Recording parameters (for example, amplifier gain or recording onset delay) were adjusted in order to provide the clearest ECAP response. Once a clear ECAP response was obtained in this condition, the location of the masker electrode was varied in a staggered fashion and a repeat NRT recording obtained for each masker location. For example, a recording series might include masker electrodes in a sequence such as: 11, 10, 12, 9, 13, 8, 14, 7, and so on until the basal and apical ends of the array were reached. This staggering was used to mitigate against any fatigue effects that might occur in a recording sequence. Following each recording series, the NRT responses were analysed offline to measure the ECAP response amplitude for each probe/masker combination, as described above (see figure 8.2).

### 8.2.4 Listeners

Twenty four experienced adult users of the Nucleus CI24 or Nucleus 'Freedom' cochlear implant were tested. Demographic and other details are shown in table 8.1. All had normal, full electrode insertions. Eight listeners (those listed with an L prefix) had also participated in one or more of the other experiments reported in this thesis. The other listeners were selected at random from the patient population at the Midlands Adult cochlear implant programme, University Hospital Birmingham NHS Trust.

**Table 8.1**

Demographic and clinical details of implant users tested.

CI24RCS and CI24RCA = Nucleus 'Contour' electrode array

CI24M = Nucleus straight electrode array; CI24RE = Nucleus 'Freedom' electrode array

BKB = Open-set speech recognition in quiet, BKB sentences presented in the sound field at 70 dBA

Prefixes of notations for implant users: S = straight array; C = contour array; F = Freedom implant; L prefix indicates listeners from other thesis experiments

| Implant user | Age | Aetiology of profound deafness | Duration of profound deafness (years) | Implant device | Open-set speech recognition (BKB sentences) |
|---|---|---|---|---|---|
| L1 | 42 | Meningitis | 6 | CI24RCS | 80 |
| L2 | 47 | Otosclerosis | 5 | CI24M | 78 |
| L5 | 57 | Progressive | 5 | CI24M | 51 |
| L6 | 24 | Autoimmune | 2 | CI24M | 97 |
| L8 | 50 | Head trauma | 24 | CI24RCS | 76 |
| L11 | 41 | Meningitis | 30 | CI24RCS | 82 |
| L13 | 50 | Progressive | 2 | CI24RCA | 98 |
| L14 | 24 | Measles | 1 | CI24RCS | 87 |
| S2 | 56 | Idiopathic | 50 | CI24M | 50 |
| S6 | 55 | Progressive | 6 | CI24M | 97 |
| S7 | 86 | Progressive | 2 | CI24M | 88 |
| C1 | 75 | Progressive | 4 | CI24RCA | 16 |
| C2 | 35 | Congenital | 35 | CI24RCA | 0 |
| C3 | 72 | Idiopathic | 5 | CI24RCA | 10 |
| C4 | 22 | Congenital | 22 | CI24RCA | 21 |
| C5 | 66 | Progressive | 5 | CI24RCA | 99 |
| C6 | 79 | Idiopathic | 3 | CI24RCA | 34 |
| C7 | 61 | Meningitis | 2 | CI24RCA | 100 |
| C8 | 53 | Waardenburg's syndrome | 10 | CI24RCA | 82 |
| C13 | 19 | Idiopathic | 10 | CI24RCS | 79 |
| C14 | 75 | Progressive | 10 | CI24RCA | 80 |
| F3 | 25 | Progressive | 10 | CI24RE | 83 |
| F5 | 23 | Progressive | 2 | CI24RE | 2 |
| F6 | 70 | Autoimmune | 7 | CI24RE | 7 |

## 8.3 Results

### 8.3.1. Group results

Mean results for all 24 listeners are shown in figure 8.4, showing mean ECAP amplitudes for each location of masker electrode. Large standard errors indicate the considerable variability in the results across individuals. Also, it should be noted that the number of data points contributing to each mean shown is variable, because ECAP responses were not recorded for all masker electrode locations for all listeners owing to occasional difficulties in obtaining a clear response. ECAP amplitudes were frequently zero or close to zero at electrode locations towards the basal and apical extremes of the electrode array. The largest mean ECAP response was found with a masker electrode located at electrode 12. This is somewhat unexpected as the probe electrode was always on electrode 11 and so the peak ECAP amplitude would have been expected with the masker electrode also on e11; however, the difference in mean amplitude (4.19 µV) between electrodes 11 and 12 was small and not significant (p=0.457, paired t-test). Location of the masker on the same electrode as that used for recording (e13) frequently created a large stimulus artefact which prevented recording of a response. A response was recordable in this particular recording condition in only 10 cases (hence the large standard error for that point), but the mean ECAP amplitude has been included for the purposes of illustration. The general trend towards reduced ECAP amplitudes with increasing masker-probe separation in both basal and apical directions is apparent. Also, an overall trend towards slightly higher ECAP amplitudes with masker electrode locations towards the apical end of the array (higher electrode numbers) than towards the basal end (lower electrode numbers) is apparent: the mean amplitude for a masker location on electrode 22 was 58.7 μV compared with 7.3 μV on electrode 1 (p=0.016, t-test). This suggests a trend towards greater spread of excitation in the apical direction than in the basal direction. This was mainly caused by 5 listeners who showed asymmetrical profiles, as discussed below. A within-subjects, two-factor ANOVA showed a significant main effect of electrode separation (this included 11 separations from electrode 11, including a separation of zero, in each direction): F(10,50)=3.596, p=0.001. There was no significant effect of direction (of electrode separation, i.e. basal vs. apical): F(1,5)=1.883, p=0.228, and no significant interaction between electrode separation and direction of separation: F(10,50)=1.602, p=0.133.

### 8.3.2 Individual results

Individual results showing NRT spread of excitation profiles for all listeners are shown in figures 8.5 and 8.6. In each case the abscissa shows the electrode number for the masker electrode; the ordinate shows the ECAP (Evoked Compound Action Potential) response amplitude. Considerable variability in both the size and pattern of responses is apparent. The maximum ECAP response amplitude ranged from 19 to 638 μV (median = 135 μV, standard deviation = 143 μV), and was most frequently recorded for a masker located on electrode 11 (10 cases) or on electrode 12 (9 cases). In 2 cases, maximum amplitude was recorded with the masker on e10 and in 3 cases with the masker on e13. Spread of excitation was generally fairly broad, with evidence of narrowly defined excitation only apparent in a small number of listeners, for example L1 and F3. Nineteen listeners showed a roughly symmetrical profile each side of electrode 11. The remaining 5 (L6, L11, C4, C5 and C8) produced asymmetrical patterns, with raised ECAP amplitudes towards the apical end of the electrode array.

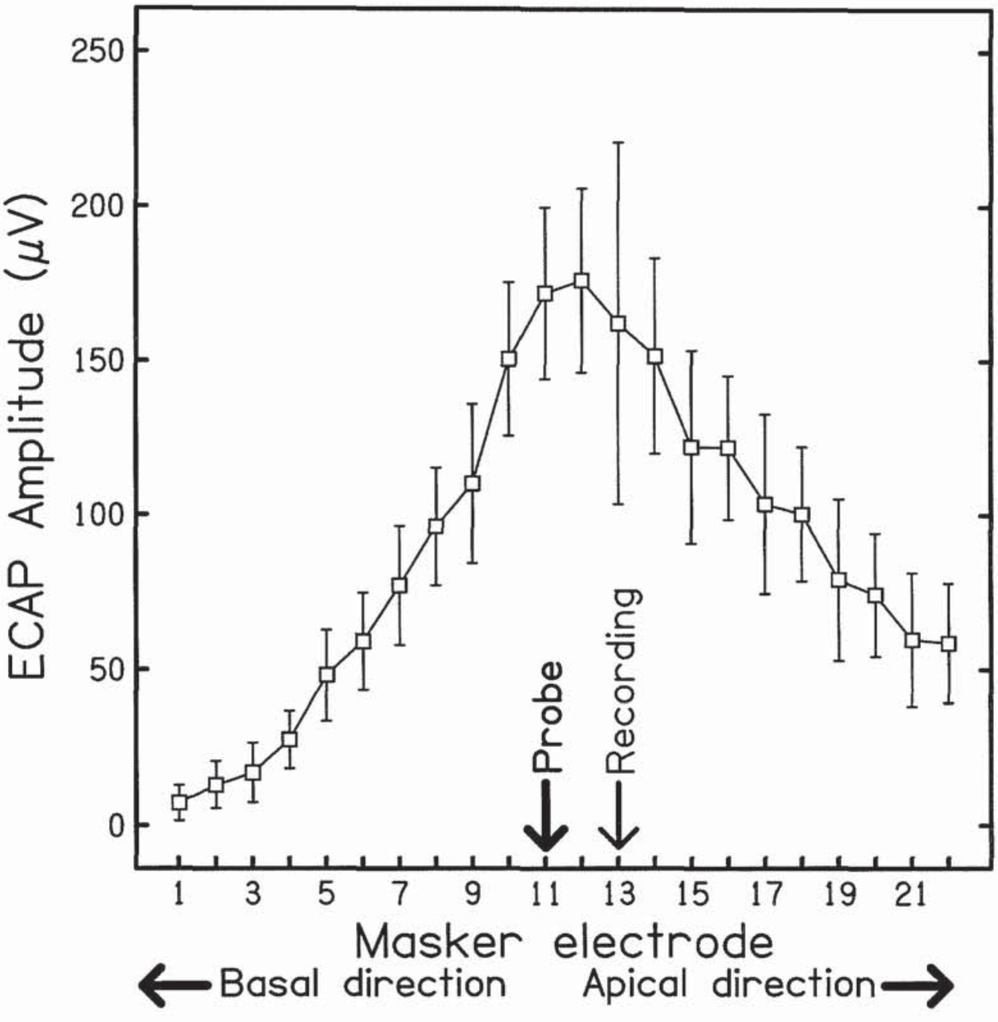
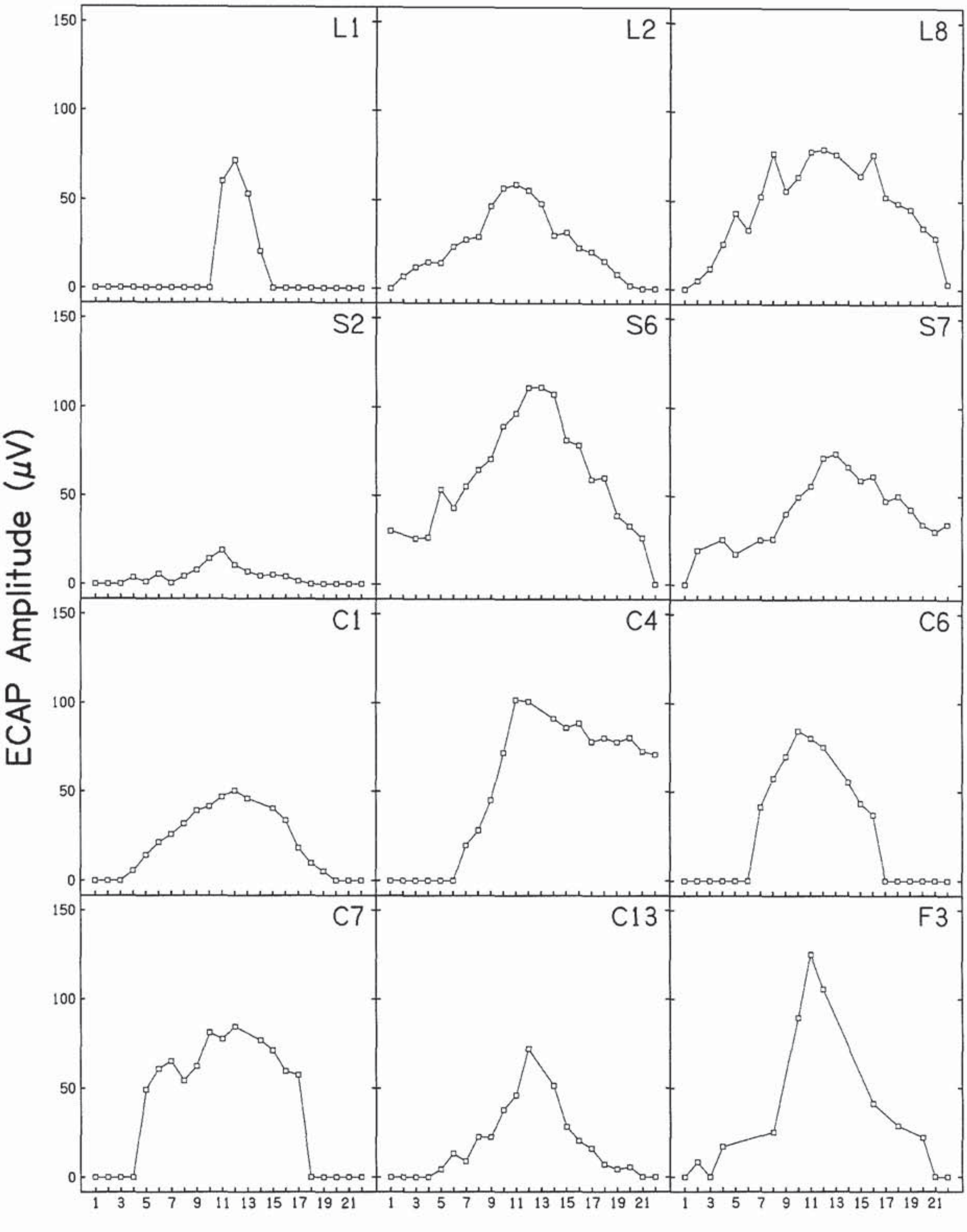
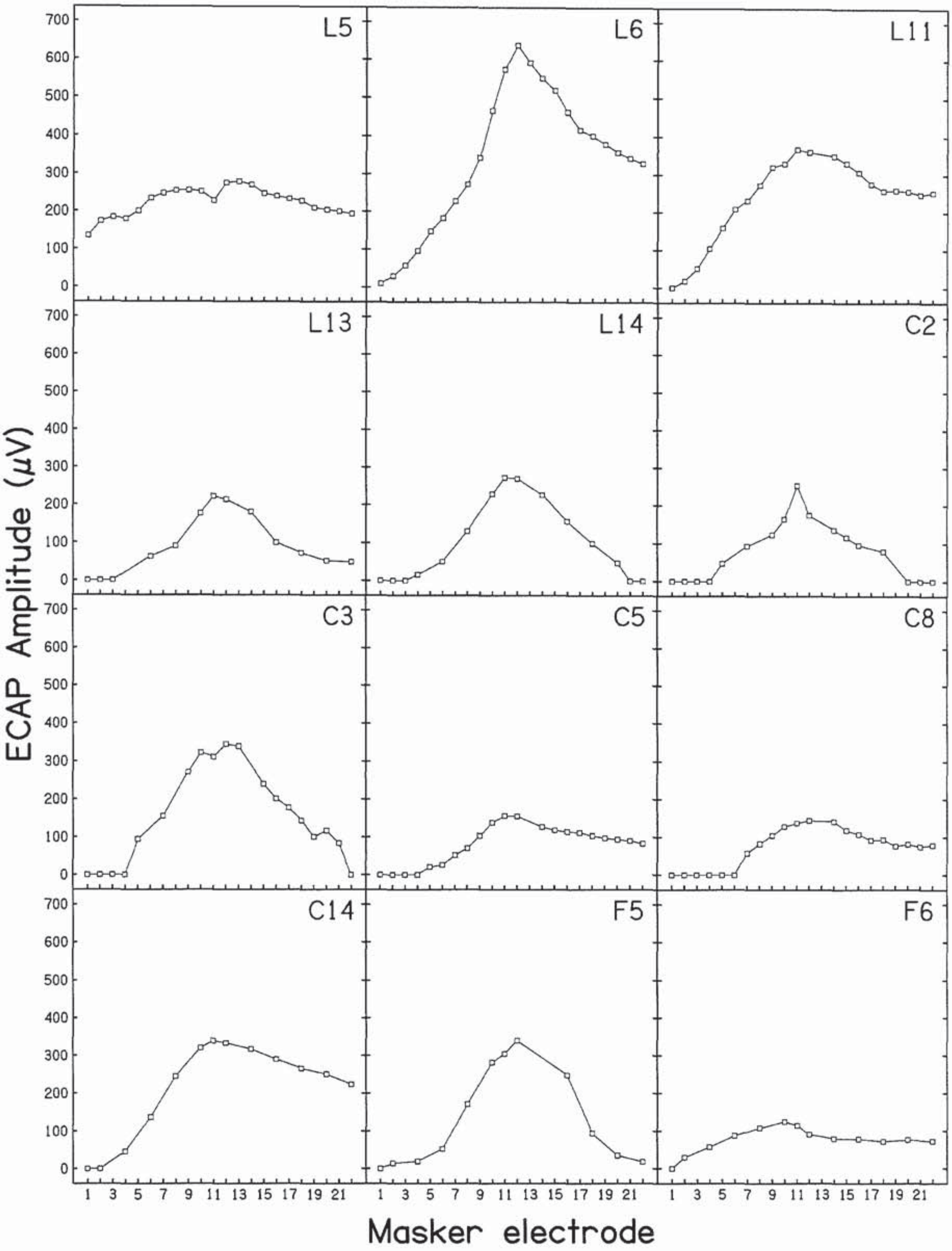


**Figure 8.4.** Mean ECAP peak-to-peak amplitudes averaged across all listeners. The abscissa indicates the location in the array of the masker electrode. The ordinate shows the peak-to-peak amplitude of the ECAP response in μV. The probe electrode was always on electrode 11, as shown by the left vertical arrow. The recording electrode was number 13, as shown by the right vertical arrow. Vertical lines show standard errors for each mean. N.B.: Where data were missing for specific masker electrode locations, means were calculated only from the available individual data points.


L1
L2
L8
S2
S6
S7
C1
C4
C6
C7
C13
F3
ECAP Amplitude (μV)
150
100
50
0
Masker electrode
1
3
5
7
9
11
13
15
17
19
21



**Figures 8.5 and 8.6.** Individual results for all 24 implant users. For the purposes of illustration, the listeners have been divided into two groups of 12. Figure 8.5 includes those with relatively smaller ECAP amplitudes, and the ordinate scale on this panel has been adjusted to show a maximum amplitude

of 150 µV. Figure 8.6 shows results for the other 12 listeners, who had generally larger ECAP amplitudes; the ordinate is adjusted to a maximum of 700 µV. The x-axes indicate the location of the masker electrode; the probe electrode was always on electrode 11. Within each figure, the order in which listeners' data is shown corresponds to Table 8.1. Where data values were missing, no symbol is shown and neighbouring values were interpolated.

### 8.3.3. Calculation of ECAP profile widths

To provide a standardised measure of spread of excitation, the profile width at 75% of maximum amplitude was calculated for each listener. To do this, the maximum amplitude was first calculated. From this, the point equal to 75% of maximum was calculated, in µV, and the width of the ECAP profile at that point was measured, expressed in terms of number of electrodes. The 75% point was used as it was found to be the lowest point at which full profile widths could be measured for all listeners; also, it allowed comparison with previously published data (e.g. Hughes and Abbas, 2006). In cases where an ECAP response was not recorded or recordable for a particular location of the masker, a linear interpolation between data points for neighbouring electrodes was used; the width measure was derived from this plot. A similar method was used by Hughes and Abbas (2006). An illustration of this method is shown in figure 8.7. The median profile width was 5.80 electrodes; standard deviation = 3.2 electrodes; maximum = 1.5; minimum = 14.0. Although a wide range of profile widths was found, just over half (13) of the listeners produced widths in the range between 4 and 7 electrodes.

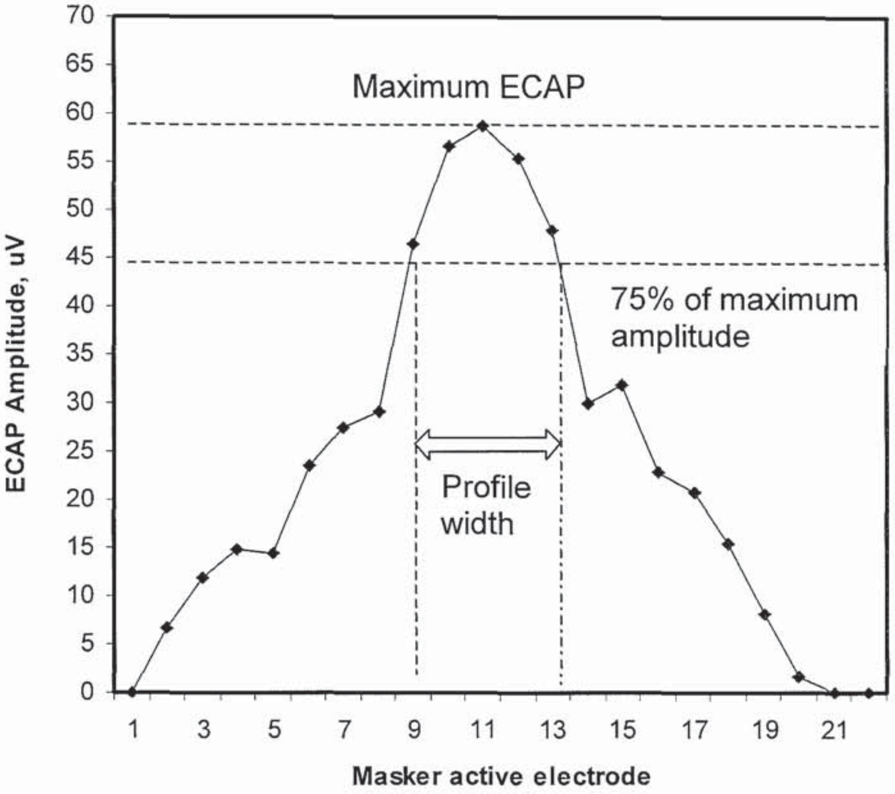


**Figure 8.7**. Method used for estimation of ECAP profile width. The abscissa indicates the location of the masker electrode. The ordinate shows the peak-to-peak amplitude of the ECAP response. The probe electrode was always electrode 11. The upper horizontal dashed line shows the maximum ECAP amplitude. The lower dashed line shows the point at 75% of maximum amplitude. The profile width is measured at this point in units of electrode number.

### 8.4 Correlations between NRT/ECAP data and perceptual results

As discussed above, ECAP data measured using the VMFR method provide an objective estimate of the spread of excitation in the cochlea. The results obtained from this sample of 24 implant users demonstrate the variability in both ECAP amplitudes and the shape and width of ECAP profiles, expressed in terms of numbers of electrodes. If spread of excitation measured in this way does relate to psychophysical measures, where channel discrimination is expected to be an important factor, then correlations should be apparent between the ECAP profile data and the results for those other tasks. In order to investigate whether any such correlations existed between the ECAP data recorded here and either speech recognition performance or results from other psychophysical tasks (described in chapters 3 and 4), the results were analysed in comparison with those other measures. The most straightforward comparison was with results of experiments 1 and 2, concerned with reported segregation and channel discrimination, as they would be expected to relate most closely to spread of excitation. Comparison with results from the other experiments in this thesis was considered too complex and so was not pursued.

### 8.4.1 Comparison with speech recognition performance

Firstly, the relationship between the ECAP profile data and speech recognition performance using the implant was examined. Figure 8.8 shows the relationship between the ECAP profile width and open-set speech recognition scores for all 24 listeners. The latter were measured in the sound field using the implant users' own speech processors at a 70 dB(A) presentation level in a sound-treated test room. Speech materials were unknown BKB sentences, scored on key-words correctly reported. Listeners were allowed to choose their preferred listening program and set their sensitivity or volume as desired. Fifteen out of the 24 listeners achieved open-set speech recognition above 50% correct. No significant correlation between the two measures was found; indeed, the $r^2$ value was close to zero (0.0005, $p>0.5$).

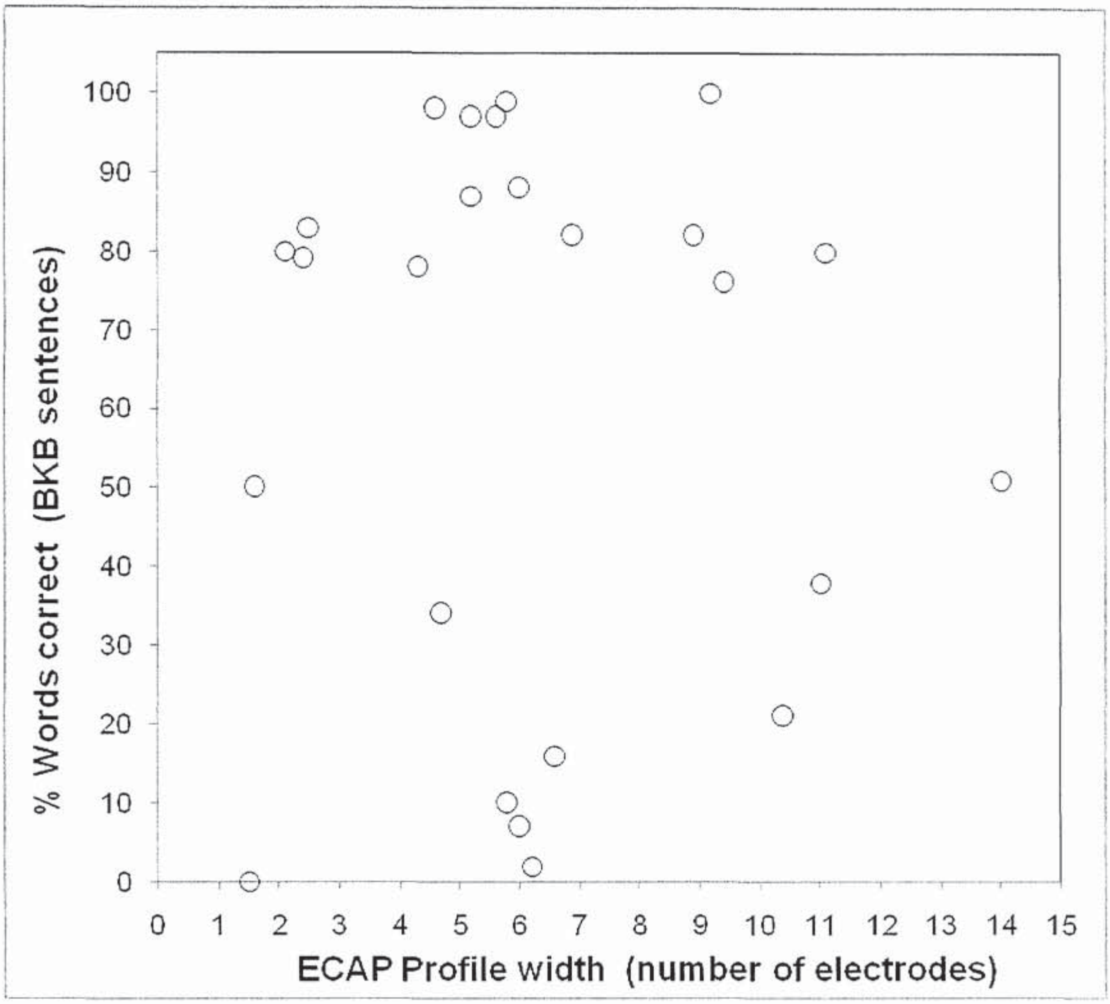


**Figure 8.8**. Scatter plot showing relationship between the ECAP profile width in number of electrodes vs. open-set speech recognition scores obtained with BKB sentences presented at 70 dB(A) in the sound field; listeners used their normal, preferred speech processing strategy and volume/sensitivity settings.

### 8.4.2 Comparison with results from experiment 1 (reported segregation)

Secondly, the ECAP spread of excitation data were analysed in relation to the results of experiments 1 and 2 (described in detail in chapters 3 and 4). The first of these gave a measure of reported segregation in a task with sequences of alternating ABA...ABA tones, where tone A was always presented on electrode 11 and tone B was presented on another electrode in the array. If *greater* reported segregation was related to *greater* discriminability and perceptual differences between electrodes, then this should also be related to *lower* spread of excitation, i.e. a smaller ECAP profile width. Four listeners (L1, L2, L5, L6) for whom ECAP data were available also participated in experiment 1 (see chapter 3 for details). In order to allow comparison between their ECAP profiles and the results of experiment 1, the data were transformed as follows:

1. ECAP amplitudes were normalised relative to the maximum recorded amplitude, so that the maximum was a value of 1. For example, for a maximum amplitude of 250 μV, an ECAP amplitude of 100 μV would be transformed to: 100/250=0.4
2. Reported segregation scores (averaged for TRTs of 100 and 150 ms - the values tested in all listeners) were transformed using the formula: $x = 1-(y/100)$, where $x$ is the transformed score and $y$ is the original score. N.B: the result of this transformation is that *higher* numbers indicate *lower* reported segregation. These values may therefore be considered as equivalent to reported *integration (scale = 0-1)*.

The transformed data are shown for individual listeners in figure 8.9. One listener (L2) demonstrated a close similarity between the profile obtained from ECAP amplitude data and her reported integration scores; i.e., *larger* ECAP amplitudes (resulting from a greater degree of overlapping stimulation with electrode 11) were closely related to a *higher* degree of reported integration and vice-versa. One other listener (L5) showed a broad spread of excitation profile in his ECAP data, corresponding to a similarly broad pattern for his (transformed) reported integration scores. The comparison between the two measures for the other two listeners is somewhat less clear. One (L1) showed a very similar slope in the profile of both measures towards the apical end of the electrode array (higher electrode numbers), but great dissimilarity in the basal direction. L6 displayed very sharply 'tuned' reported integration centred around electrode 11 and a similarly shaped profile for his ECAP amplitudes (although the peaks of each profile were at slightly different positions along the electrode array). Although he showed a strong tendency towards reported segregation (giving rise to a *lower* peak in the transformed data shown in the graph), nonetheless the correlations between his (transformed) reported segregation scores and normalised ECAP amplitudes were strong and significant (see below).

Correlations were calculated between the calculated ECAP amplitude data for each electrode and reported integration results of experiment 1, transformed as described above (excluding electrode 11, for which there was no corresponding case in the segregation task). The data corresponding to separations from electrode 11 in each direction along the array were analysed separately. The resulting correlations are shown in table 8.2. All 7 correlations calculated between ECAP amplitudes and transformed reported segregation were highly significant (in both basal and apical directions). The eighth was not calculated (L1, basal direction) owing to insufficient non-zero ECAP data.
The overall correlation was strong and significant in the basal direction (pooled data from three listeners, but not including L1 as there was insufficient ECAP data available), accounting for around 75% of the variance. In the apical direction, the overall correlation (pooled data from all four listeners) was

small but significant, accounting for about 10% of the variance. Figure 8.10 shows scatter plots of ECAP amplitude against reported integration scores for each direction of electrode separation.

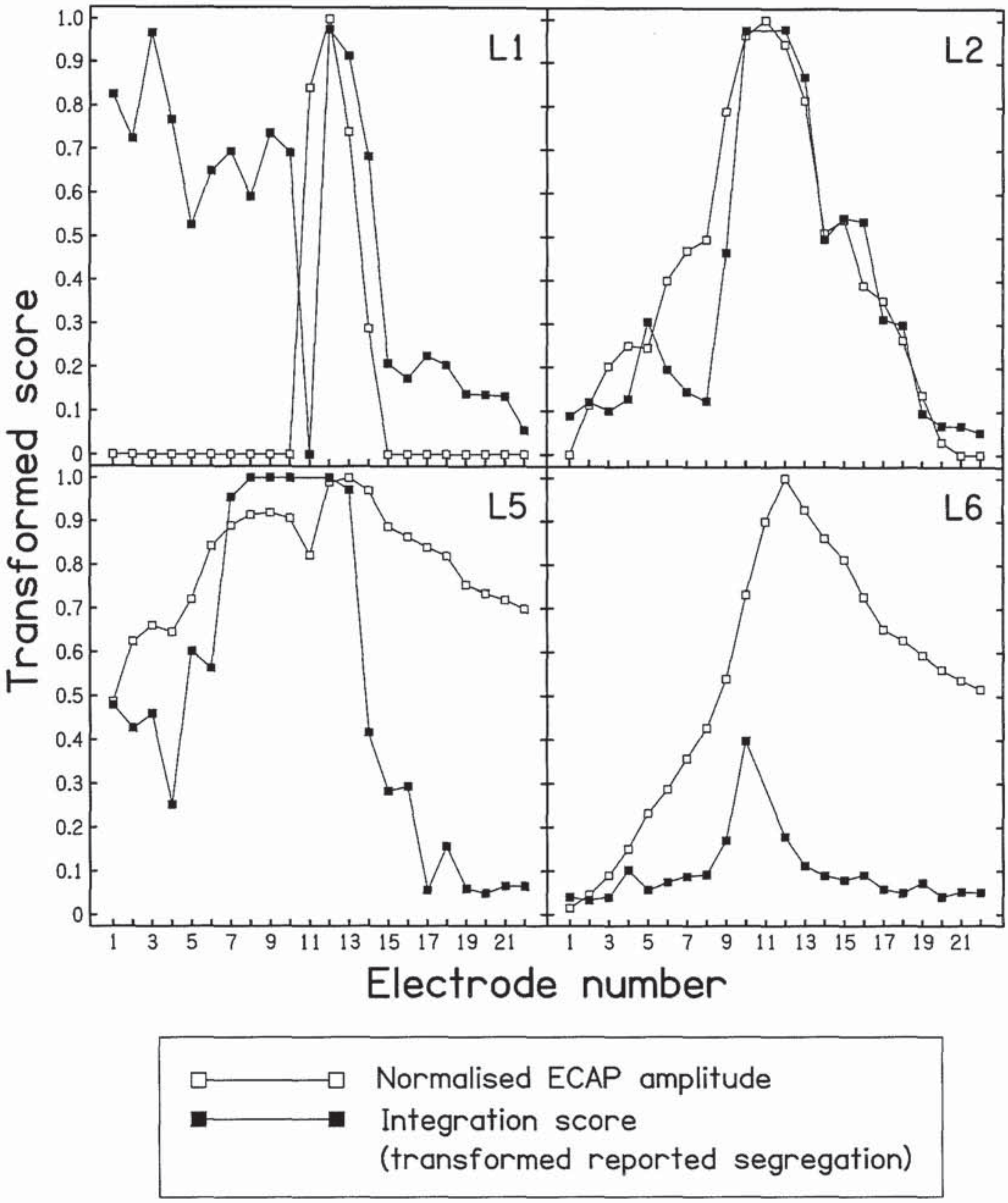


**Figure 8.9**. Normalised ECAP amplitudes and integration scores (transformed reported segregation) for four listeners. The abscissa label (Electrode number) refers to: a) the masker electrode for the ECAP data and b) the

electrode number for tone B (see experiment 1, chapter 3). The ordinate shows the transformed segregation scores. N.B: Reported segregation scores have been transformed into a scale from 0 to 1, and inverted for the purposes of illustration and comparison with the ECAP data, so that higher values (of the derived integration score) relate to *lower* reported segregation. ECAP amplitudes have been normalised so that the maximum amplitude is 1.0.

**Table 8.2**. Correlations (adjusted r squared values) between normalised ECAP amplitudes and transformed reported integration scores. Separate correlations were calculated for each half of the electrode array (i.e.: basal, electrodes 1 to 10, and apical, electrodes 12 to 22). Correlations that are statistically significant ($p<0.05$) are shown in bold text. Slopes and intercepts of regression lines fitting the data are also shown.

| Listener | Direction of electrode separation | ECAP amplitude vs. reported integration | | | |
|---|---|---|---|---|---|
| | | Slope | Intercept | Adjusted $r^2$ | p values |
| L1 | Basal | *Insufficient ECAP data available* | *Insufficient ECAP data available* | *Insufficient ECAP data available* | *Insufficient ECAP data available* |
| | Apical | 0.91 | 0.18 | **0.903** | **<0.001** |
| L2 | Basal | 0.76 | -0.03 | **0.660** | **0.001** |
| | Apical | 1.00 | 0.03 | **0.970** | **<0.001** |
| L5 | Basal | 1.59 | -0.54 | **0.688** | **0.002** |
| | Apical | 2.79 | -2.04 | **0.725** | **0.001** |
| L6 | Basal | 0.41 | -0.01 | **0.703** | **0.002** |
| | Apical | 0.21 | -0.07 | **0.753** | **<0.001** |
| Overall | Basal (excluding L1) | 0.95 | -0.11 | **0.753** | **<0.001** |
| | Apical | 0.31 | 0.12 | **0.116** | **0.013** |

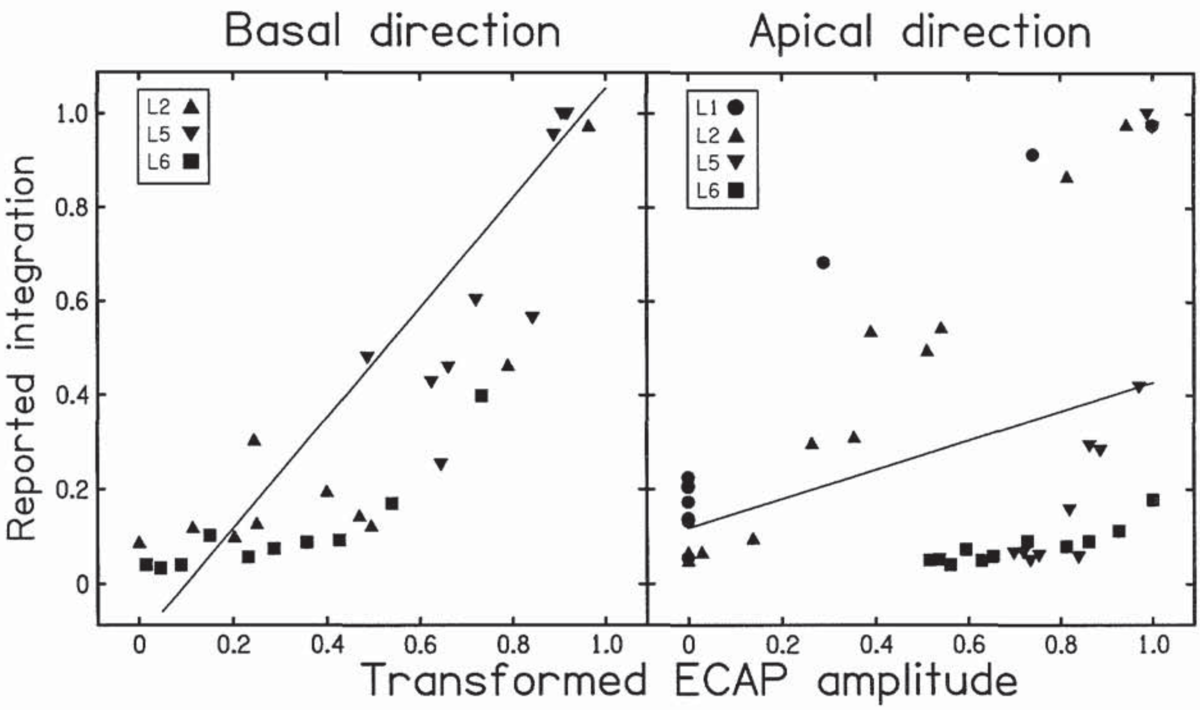


**Figure 8.10**. Scatter plots between normalised ECAP amplitudes and transformed reported integration scores. Different symbols are used for each individual listener's data. Linear regression lines fitting the data are shown.

### 8.4.3 Comparison with results of experiment 2 (pitch discrimination)

Eight listeners for whom ECAP data were recorded participated in the pitch discrimination task as used in experiment 2 (see chapter 4 for full details of task and procedure). Their ECAP amplitude data were transformed as described above. Their pitch ranking and confidence rating scores were firstly transformed using the same method as described in chapter 4, i.e.:

(i) the percentage of responses where each electrode was reported as higher in pitch than electrode 11 was converted into an absolute difference from chance (50%) and re-scaled from 0 to 100 (i.e., absolute difference x 2), thus giving a measure of channel discrimination;

(ii) The confidence ratings were re-scaled from the 1 to 5 range used by the listeners into a 0 to 100 scale (i.e., 25 x (rating – 1)).

Secondly, each of these measures was *further* transformed to allow comparison with the transformed ECAP data using the same formula as described above for the reported segregation data, i.e. $x = 1-(y/100)$, where x is the result of this second transformation, and y is the previously transformed score. The transformed scores are shown in figures 8.11 and 8.12. N.B: these transformations of the data result in *higher* numbers indicating *poorer* pitch discrimination of each electrode from electrode 11 and *lower* confidence in the pitch ranking judgements made. Unlike for the comparison with reported segregation, it should be noted that data for electrode 11 are included because they were available for both experiments.

Inspection of the individual paired profiles shows that in the comparison with the channel discrimination data (figure 8.11), a close correspondence with normalised ECAP amplitudes is not apparent for most listeners, although both curves generally peak around electrode 11 (or close to it) and decline with increasing electrode separation in either basal or apical directions.

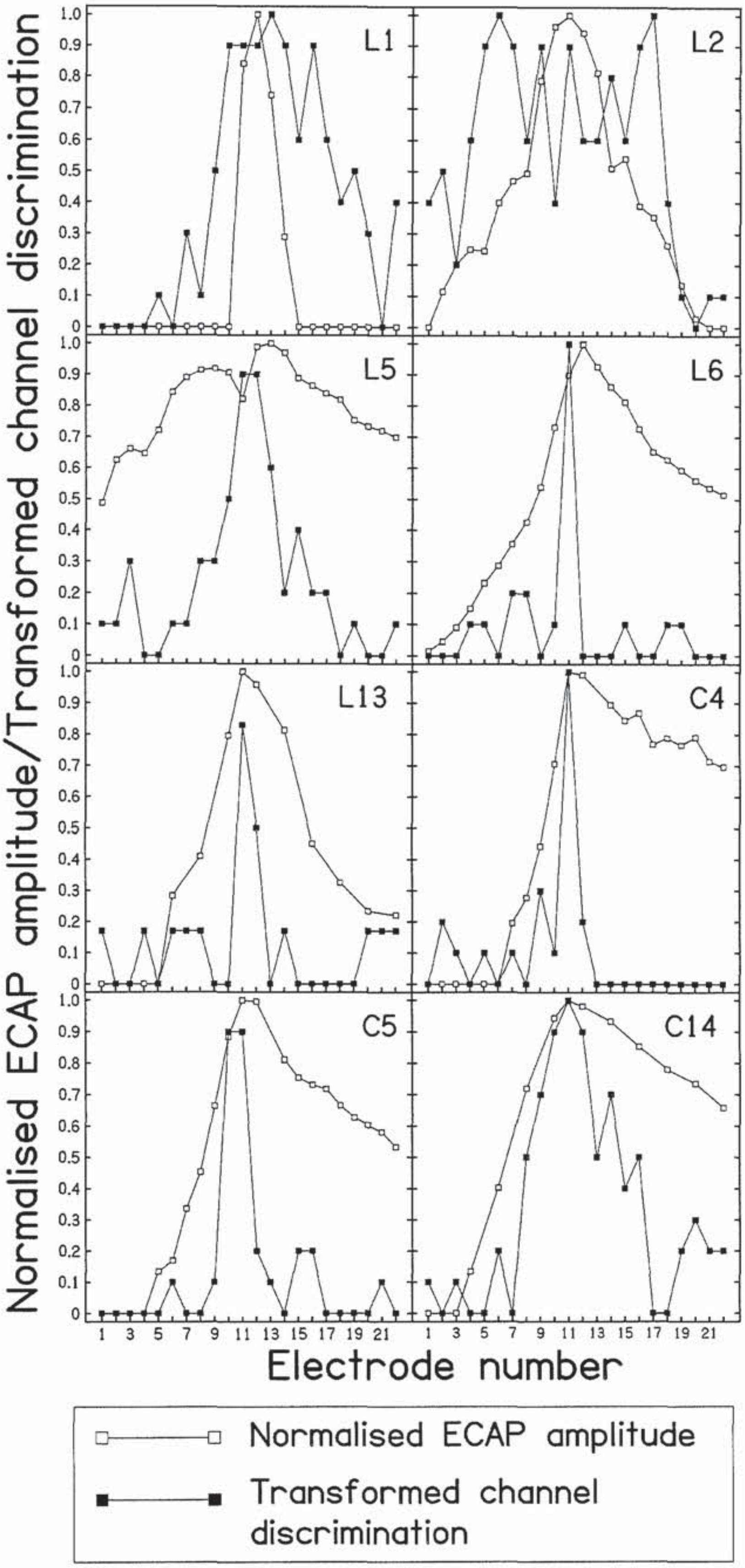


**Figure 8.11**. Normalised ECAP amplitudes and transformed channel discrimination data: individual results for 8 listeners

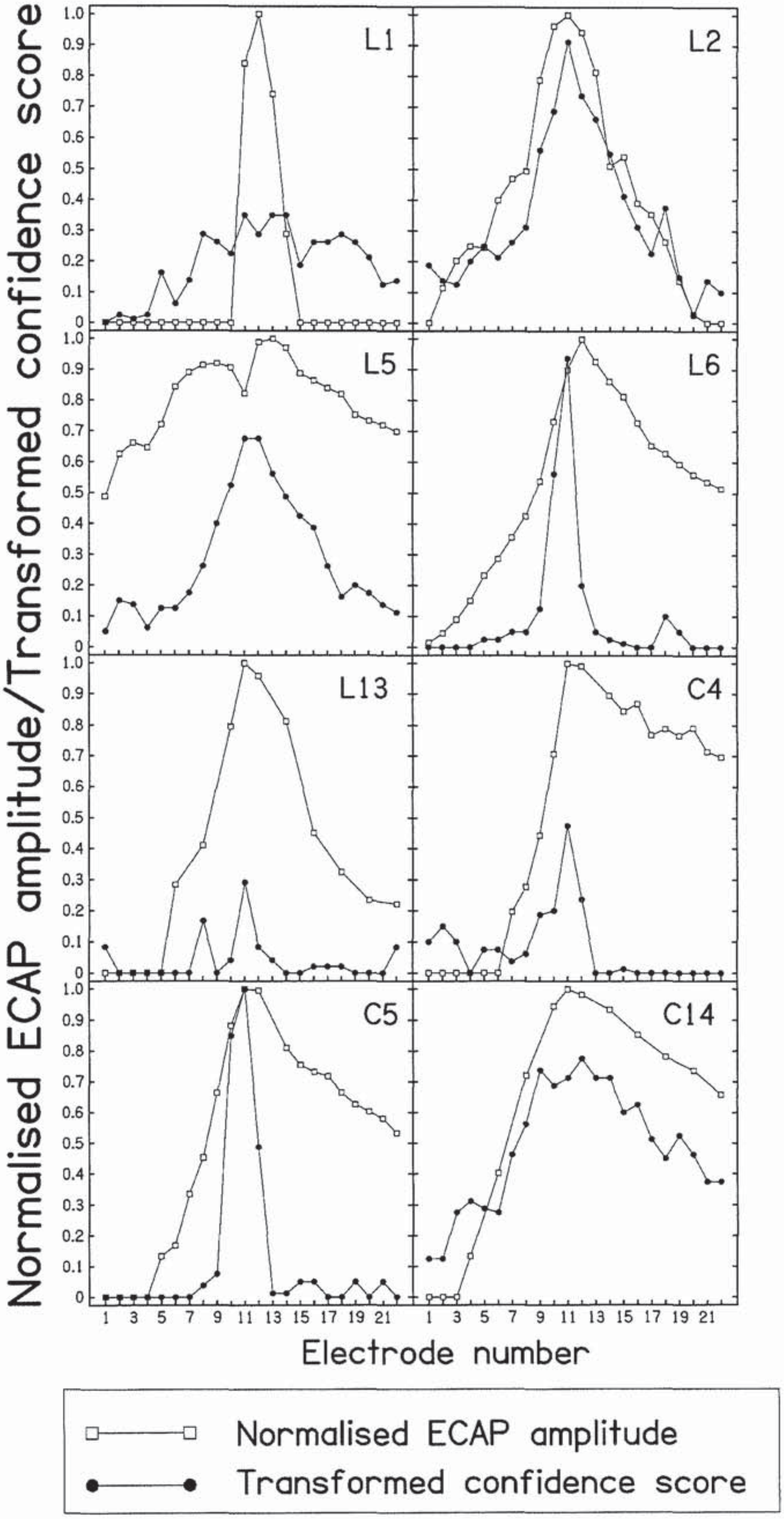


**Figure 8.12**. Normalised ECAP amplitudes and transformed confidence scores: individual results for 8 listeners.

A somewhat closer correspondence between (transformed) reported confidence scores and normalised ECAP amplitudes (figure 8.12) is apparent. For example, a close match between the two profiles is seen for listener L2; the match is slightly less close but still reasonable for C14. It should be noted that the transformed channel discrimination and confidence scores were not normalised (as the ECAP amplitudes were) relative to a peak value of 1. Correlations were computed for each listener, and for the group data, between the calculated ECAP amplitude data for each electrode and the corresponding transformed channel discrimination and confidence scores, one for each direction in the electrode array (see table 8.3). In both cases, 12 out of the set of 16 correlations for individual listeners were significant ($p<0.05$). Overall correlations (pooled data from all listeners) between normalised ECAP amplitudes and transformed channel discrimination were positive and significant in the basal direction, but not in the apical direction, accounting for about 30% and 1% of the variance respectively. Overall correlations were positive and significant in both directions between normalised ECAP amplitudes and transformed confidence scores, accounting for about 40% (basal direction) and 10% (apical direction) of the variance. Figure 8.13 shows scatter plots illustrating the relationship between normalised ECAP amplitudes and a) transformed channel discrimination and b) transformed confidence scores for all eight listeners; the stronger relationship between normalised ECAP amplitudes and (transformed) confidence scores (upper two panels) than with transformed channel discrimination (lower two panels) is apparent.

**Table 8.3**. Correlations (corrected r squared) between normalised ECAP amplitudes and (a) transformed channel discrimination and (b) confidence scores. Separate correlations were calculated for each half of the electrode array (i.e.: basal: electrodes 1 to 11, and apical: electrodes 11 to 22). Correlations that were statistically significant ($p<0.05$) are shown in bold text. Slopes and intercepts for linear regression lines fitting the data are also shown.

(a)

| Listener | Direction of electrode separation | ECAP amplitude vs. transformed channel discrimination | | | |
|---|---|---|---|---|---|
| | | Slope | Intercept | $r^2$ | p values |
| L1 | Basal | 0.84 | 0.19 | **0.291** | **0.050** |
| | Apical | 0.54 | 0.48 | **0.396** | **0.017** |
| L2 | Basal | 0.27 | 0.54 | 0.019 | 0.303 |
| | Apical | 0.70 | 0.22 | **0.424** | **0.013** |
| L5 | Basal | 0.72 | -0.31 | 0.065 | 0.225 |
| | Apical | 1.94 | -1.33 | 0.319 | **0.033** |
| L6 | Basal | 0.71 | -0.09 | 0.344 | **0.027** |
| | Apical | 0.51 | -0.26 | 0.000 | 0.340 |
| L13 | Basal | 0.45 | 0.04 | 0.343 | 0.057 |
| | Apical | 0.68 | -0.13 | **0.502** | **0.045** |
| C4 | Basal | 0.65 | 0.02 | **0.553** | **0.003** |
| | Apical | 1.97 | -1.52 | **0.375** | **0.027** |
| C5 | Basal | 0.81 | -0.09 | **0.675** | **0.001** |
| | Apical | 1.86 | -0.72 | **0.426** | **0.018** |
| C14 | Basal | 0.89 | -0.01 | **0.903** | **<0.001** |
| | Apical | 2.56 | -1.65 | **0.759** | **0.007** |
| Overall | Basal | 0.53 | 0.08 | **0.312** | **<0.001** |
| | Apical | 0.14 | 0.24 | 0.010 | 0.354 |

(b)

| Listener | Direction of electrode separation | ECAP amplitude vs. transformed confidence scores | | | |
|---|---|---|---|---|---|
| | | Slope | Intercept | $r^2$ | p values |
| L1 | Basal | 0.27 | 0.12 | 0.229 | 0.077 |
| | Apical | 0.12 | 0.23 | **0.328** | **0.030** |
| L2 | Basal | 0.71 | 0.03 | **0.854** | **<0.001** |
| | Apical | 0.76 | 0.07 | **0.923** | **<0.001** |
| L5 | Basal | 0.83 | -0.39 | **0.291** | **0.050** |
| | Apical | 1.59 | -0.98 | **0.594** | **0.002** |
| L6 | Basal | 0.93 | -0.16 | **0.743** | **<0.001** |
| | Apical | 0.69 | -0.39 | 0.111 | 0.155 |
| L13 | Basal | 0.19 | 0.01 | **0.467** | **0.025** |
| | Apical | 0.17 | -0.03 | 0.322 | 0.184 |
| C4 | Basal | 0.20 | 0.14 | 0.015 | 0.304 |
| | Apical | 1.16 | -0.89 | **0.544** | **0.006** |
| C5 | Basal | 0.85 | -0.1 | **0.685** | **0.001** |
| | Apical | 1.61 | -1.02 | **0.599** | **0.003** |
| C14 | Basal | 0.53 | 0.17 | **0.904** | **<0.001** |
| | Apical | 1.17 | -0.4 | **0.926** | **<0.001** |
| Overall | Basal | 0.45 | 0.05 | **0.428** | **<0.001** |
| | Apical | 0.26 | 0.07 | **0.093** | **0.003** |

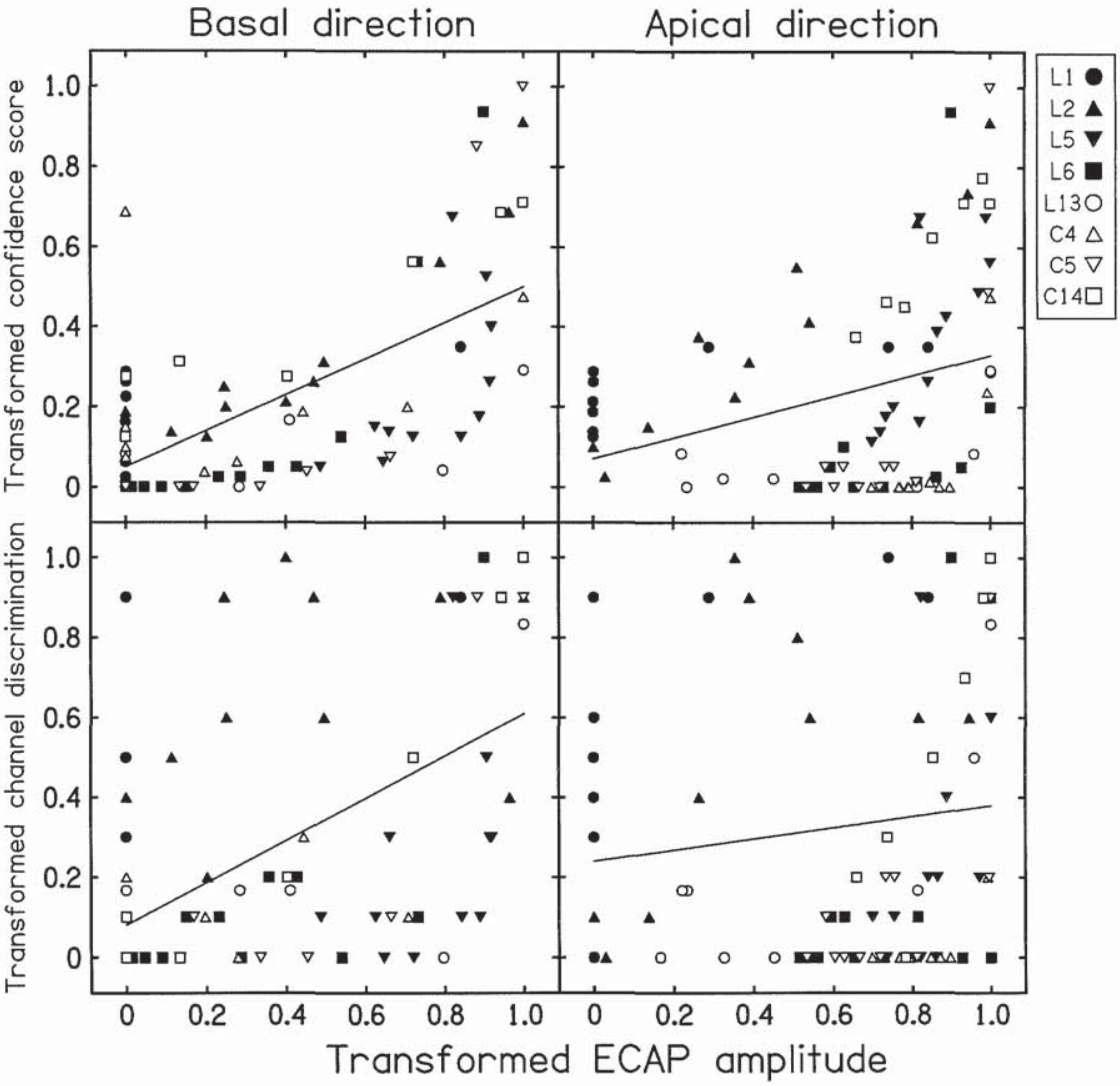


**Figure 8.13**. Scatter plots between transformed channel discrimination, transformed confidence scores and normalised ECAP amplitudes for individual listeners. Different symbols are used for each individual listener's data. Linear regression lines fitting the data are shown.

## 8.5 Discussion

The results described here suggest that the Variable Masker, Fixed Recording location ECAP measure (VMFR method) is a relatively quick and easy way to estimate spread of excitation in the cochlea resulting from electrical stimulation. Previous studies (e.g. Cohen et al., 2003) have demonstrated that the excitation profiles obtained using this method are in good agreement with results of psychophysical forward-masking studies. Although there was considerable variability in the absolute amplitude of the ECAP responses recorded, the shapes of the profiles obtained resulting from spatial separation of the masker from the probe electrode are similar to those reported in previous studies (e.g. Cohen et al., 2003; Hughes and Abbas, 2006). In most cases, maximum ECAP amplitudes were obtained when masker and probe coincided on the same electrode (e11 in this experiment). ECAP amplitudes then fell with varying steepness and in most cases in a monotonic fashion as the separation between masker and probe increased in either the basal or apical direction. Some asymmetry in the ECAP profile on each side of e11 was observed for some listeners (5 out of 24) and in the mean data, i.e. the slope of decline in ECAP amplitude as the masker was separated from probe was somewhat shallower in the apical direction than in the basal direction. However, no significant effect of direction (apical vs basal) was found in the ANOVA. A similar finding was reported by Cohen et al. (2003); they suggested that an asymmetry in the transmission of the stimulation field along the scala tympani would be expected due to the tapered spiral structure of the cochlea. If this hypothesis is correct, there may be anatomical reasons for the somewhat greater spread of excitation towards the apex than towards the base of the cochlea that was seen for some (though by no means all) listeners. Hughes and Abbas (2006) reported a similar asymmetry in some of their subjects; they hypothesised that it may have been due to uneven patterns of nerve survival or non-uniform current spread throughout the cochlea.

One striking aspect of the individual results is the fact that an ECAP response was still recordable in many cases, although with reduced amplitude, even for large separations between probe and masker. Some overlap in the excitation patterns for the probe and masker is necessary for any ECAP to be recorded using this method, so the fact that a response was recorded with the masker on e22 in 11 out of 24 listeners is remarkable (as the probe was always on e11); this demonstrates the considerable spread in excitation that was often present. If the point at which the ECAP amplitude fell to 50% of its maximum either side of e11 (e.g. as used by Cohen et al., 2003 to measure ECAP profile width) is examined, a spread of around 11 electrodes is seen in the averaged data. Thus, even a probe/masker separation of over 5 electrodes on average resulted in only a 50% reduction in ECAP amplitude, suggesting considerable overlap in the excitation patterns of each electrode.

The relationship between the ECAP profiles and other measures yielded interesting results. Firstly, no correlation was found between ECAP profile width and speech recognition performance with the implant. This is consistent with the findings of Hughes and Abbas (2006); they also found no statistically significant correlations between the width of their ECAP masking functions and three measures of speech recognition performance in 10 implant users. However, use of such 'width' measures may not be without problems, as two profiles could have the same width but different shapes; also, this method collapses all the data into a single measure. Furthermore, it may be that open-set speech recognition using sentence materials is an insufficiently sensitive measure to identify subtle differences in performance; more specific measures, e.g. of vowel or consonant recognition, may provide more sensitivity. Such measures should avoid some of the linguistic or cognitive influences on performance that may contribute to the variance in performance with sentence materials, and would also be expected to be more specifically related to spectral resolution. Also, a number of uncontrolled variables are thought to influence speech recognition performance with a cochlear implant, for example the speech-processing strategy implemented in the speech processor, the volume/sensitivity settings used during the testing etc. These might have been sufficient to obscure any correlation.

The second comparison made was between the ECAP profile data and the results from experiment 1, in which listeners reported whether they heard a segregated percept while listening to ABA..ABA..ABA sequences of alternating tones. Tone A was on electrode 11, the same as the probe electrode used for recording the ECAP responses. Increased electrode separation between tone A and tone B (in either apical or basal directions) led to increased reported segregation; presumably, this resulted from increased perceptual differences between A and B as they were more spatially separated across the array. These differences may be assumed to result from differing neuronal populations being stimulated by tone A and tone B; thus, the same electrode separations should result in *reduced* ECAP response amplitudes with *increasing* separation. If the two measures, one psychophysical and the other electrophysiological, were indeed measuring the same underlying processes, then a close correspondence should be found between them. The fact that highly significant individual correlations between ECAP amplitudes and reported (transformed) segregation scores were found for the 4 listeners (mean $r^2$= 0.77) suggests that the degree of overlap in neuronal stimulation between two electrodes did indeed relate to the degree of perceptual difference between them. The overall correlation between these two measures (pooled data for 3 listeners) was also positive and strong in the basal direction, but somewhat weaker in the apical direction. Although data from only a small group of implant users is presented here, this finding does suggest that the NRT-ECAP method can be used to evaluate the underlying physiological basis of perceptual differences.

The relationship between the ECAP profiles and results from the electrode pitch discrimination task was less clear cut. Correlations for individual listeners between ECAP amplitudes and channel discrimination performance were in many cases small (mean $r^2$= 0.4), although 12 out of the 16 were significant. Overall correlations (pooled data from all eight listeners) were significant only in the basal direction ($r^2$=0.31); visually there was not a close correspondence between the two measures for most listeners. Hughes and Abbas (2006) found no significant correlation between the width of their ECAP profile functions and the slope of an electrode pitch ranking function in their 10 implant users.

A slightly stronger relationship was observed between the ECAP data and reported *confidence* in pitch judgements (mean $r^2 = 0.53$) and several individual correlations were somewhat higher; again, 12 out of the 16 were significant. Also, significant correlations were found for the pooled data from all eight listeners, although stronger in the basal direction ($r^2$=0.43) than in the apical direction ($r^2$=0.09). In other words, *increased* electrode separation leads to *reduced* overlap in neuronal excitation patterns between two electrodes and also to *increased* confidence reported in judgements of pitch differences between them. This may not be entirely surprising, as similar results were found in the analysis of the results of experiments 1 and 2 (see chapter 4); in that case, reported segregation correlated more strongly with confidence scores on the pitch discrimination task than with actual pitch discrimination performance. The actual pitch percepts available to cochlear implant listeners may be unclear or indistinct, and perceptual differences arising from stimulation on different electrodes may be more akin to changes in timbral brightness rather than musical pitch (see chapter 2 for further discussion). Therefore, the relationship between electrode pitch discrimination and the ECAP measure of spread of excitation may be quite a loose one. Also, although many of the comparisons discussed above produced significant correlations, inspection of the comparison plots (figures 8.9, 8.11, and 8.12) shows that in some cases the profiles for individual listeners could differ quite a lot but still produce a significant correlation. Notwithstanding the problems associated with condensing the ECAP data into a single width measure, this begs the question of how much weight should be attributed to the finding of significant correlations between the slopes of profiles derived from different tasks.

## 8.6 Conclusion and future directions

The NRT-ECAP 'VMFR' method is a relatively quick and objective way of measuring spread of excitation in the cochlea resulting from electrical stimulation and so to assess channel interaction. Previous studies had not found significant correlations between the results obtained using this method and psychophysical measures of electrode pitch ranking or speech recognition; however, they were based on a profile width measure which may not be optimal for revealing the correlations that were sought.

ECAP profiles recorded in 24 implant users showed that often the spread of excitation in the cochlea is substantial; on average, a separation of around 5 electrodes only reduced the response amplitude by 50%. No relationship between the width of the spread of excitation measures and speech recognition performance with the implant was found; this is consistent with previous reports and explicable in terms of the number of variables that are thought to influence speech recognition performance and the relative insensitivity of sentence materials to subtle performance differences.

Significant correlations between the ECAP spread of excitation profiles and (transformed) reported segregation in a task aimed at evaluating stream segregation were found in 4 listeners for whom data were available; this suggests that *reduced* overlap in areas of neuronal stimulation between two stimulated electrodes leads to *increased* reports of segregation in that task. Thus, some objective evidence of the physiological basis of the segregation task was found, although based on correlations, which of course do not prove causation. A much weaker relationship between ECAP spread of excitation and pitch discrimination performance was found; an intermediate correlation with reported *confidence* in pitch judgements suggested that reduced spread of excitation can lead to increased confidence in judgements about perceptual differences between two electrodes.

In summary, the ECAP 'VMFR' method was used to demonstrate the generally broad spread of excitation in a sample of implant users. Some evidence of a relationship between this measure and psychophysical measures of either reported segregation or channel discrimination was found, although not a strong one. Future research could explore the relationship of the ECAP-VMFR data with speech perception measures expected to be more sensitive to spectral resolution; also, measuring the correlation of ECAP data with the results of other psychophysical tasks where spectral resolution is an important variable may also be worthwhile.

# Chapter 9: Discussion

## 9.1 Introduction

Cochlear implants have proved successful clinically in providing hearing sensations to severely or profoundly deaf adults and children. Although they are unable to restore 'normal' hearing, they can in many cases provide astonishingly good hearing abilities to users. In adults, it is now expected that most implant recipients will be able to achieve at least some degree of open-set speech recognition in quiet conditions. In many cases, this allows the best implant users to carry out fluent spoken communication without lipreading cues, such as holding conversations over the telephone. However, most implant listeners experience significant difficulties and decrements in their performance when listening in noisy conditions or in the presence of competing speech. The majority of implant users are monaural listeners, because they have only one implant and so effectively remain deaf in one ear. They are therefore denied the benefits of binaural hearing that are available to normally hearing listeners, such as the ability to perform auditory scene analysis based on spatial cues.

The signal that is conveyed to cochlear implant listeners is also severely restricted by the limitations of implant design. Currently available implant electrode arrays have at most 22 intracochlear electrodes. Despite considerable efforts that have been made to improve the selectivity of neuronal excitation from individual electrodes, in reality significant spread of excitation is inevitable, so that electrodes near to each other in the array create overlapping areas of stimulation and hence channel interaction. Thus, the quality of spectral information that can be conveyed is probably significantly compromised. Also, because of the way in which speech and other sounds are coded in the pattern of pulsatile stimulation delivered to the electrodes, the perception of pitch is frequently poor. Individual harmonics of complex sounds that fall within the same frequency allocation for a channel cannot be resolved, and so the fine detail of spectral content is also lost. Furthermore, the temporal fine structure that is so important for global pitch perception is not encoded by the implant. Together, these factors contribute to the finding that most implant listeners are unable to enjoy music or even recognise melodies, unless rhythmic cues are also available.

In the light of this somewhat gloomy assessment of the listening experience of cochlear implant users, one might expect that their ability to perform auditory scene analysis, in anything like the ways that are known to occur in normally hearing listeners, would be severely limited or even non-existent. Many of the cues that allow segregation of sounds in normal hearing are unlikely to be available to implant listeners. For example, as F0 is poorly conveyed by most implant speech processing strategies, implant listeners experience particular difficulties when attempting to attend to one speaker's voice in the presence of competing speech, as they are unlikely to be able to use the fundamental as a cue for segregation. Grouping based on harmonic relations is also impossible for implant listeners, as currently available speech processing strategies do not encode sounds in a way that makes harmonic relationships apparent. Despite this, stimulation of different electrodes does give rise to perceived differences in pitch (or at least timbral brightness) that implant listeners could be expected to make use of for sequential grouping of sounds. Also, in theory they should have access to level cues, including sudden change cues, which could be used for simultaneous grouping.

Despite the problems experienced when listening in background noise, anecdotal evidence suggests that cochlear implant listeners *are* apparently able to attend to target speech and ignore interfering noise to a certain degree, provided there is a sufficiently favourable signal-to-noise ratio. This may suggest that implant listeners are capable of at least some schema-based segregation of auditory inputs; a process that requires effortful listening but which does not suggest any *automatic* stream segregation. However, when the signal-to-noise ratio is good, the summed spectrum processed and encoded by the implant will be dominated by the signal; target speech could then often be recognisable based on peaks in the spectrum etc. without the need for much in the way of grouping (schema-based or otherwise). If listening monaurally (with one implant), users are already disadvantaged in challenging listening conditions. Also, it should be remembered that in many daily situations when following speech, implant users very often combine the auditory input they receive with visual cues obtained from lipreading; their reception and understanding of spoken language is therefore frequently based on analysis of both auditory and visual inputs.

The extent to which implant listeners can really demonstrate any form of auditory scene analysis and achieve either automatic or schema-based segregation has received some attention (e.g. Chatterjee et al., 2006), but the evidence so far has been somewhat inconclusive. The aim of this thesis was to investigate more systematically and thoroughly the performance of implant listeners on tasks designed to evaluate both automatic and schema-based segregation, and both sequential and simultaneous segregation.

## 9.2 Review and discussion of results

Experiment 1 aimed to evaluate sequential stream segregation using simple, easily defined and controlled stimuli that have frequently and successfully been used to investigate these processes in normal hearing (i.e. using sequences of rapidly alternating tones). The implant listeners did demonstrate a higher incidence of *reported* segregation with increased separation between the electrodes stimulated by tones A and B in an alternating ABA..ABA sequence. However, the results did not provide convincing evidence of automatic segregation because there was no significant effect of rate of presentation on reported segregation. In contrast, van Noorden (1975) and others have shown for normally hearing listeners that more rapid presentation rates (smaller TRT) are associated with a much greater degree of stream segregation (except at large frequency separations). Whilst it is true that the dependence of reported segregation on TRT may not be as strong under conditions of neutral instructions as when listeners are asked to try and hear a single stream, nonetheless there is clear experimental evidence of this dependence under such conditions (e.g. Anstis and Saida, 1985, experiment 1).

Also, there was little (if any) sign in these implant listeners of the perceptual ambiguity that is normally found in tasks of this nature; flipping between hearing a rapid alternating tone sequence as one stream or as two is a typical characteristic of stream segregation when listening to long sequences of tones. Thus, although the results clearly showed an influence of AB electrode separation on listeners' responses, they did not provide any conclusive evidence that automatic stream segregation had in fact mediated those responses. It was proposed that the task may effectively have been a measure of channel discrimination rather than of stream segregation.

Experiment 2 employed an electrode pitch ranking task (in which listeners judged the direction of pitch change for sequential pairs of stimulated electrodes) to investigate the results of experiment 1 further; this evaluated both the pitch discrimination performance of implant listeners, as well as their reported confidence in their pitch-ranking judgements. Some correlation between pitch ranking performance and reported segregation was found, but a stronger relationship between the reported *confidence* in pitch judgements and reported segregation was demonstrated. This was consistent with the idea that the results of experiment 1 primarily reflected the channel discrimination abilities of these listeners, rather than the effects of automatic stream segregation. Whilst it is also true that this pattern of results might have been found if stream segregation had been occurring, that explanation was ruled out, because of the absence of an effect of TRT on reported segregation. In the absence of any automatic stream segregation of the stimuli for most or all listeners in experiment 1, the default response may have been based simply on their ability to discriminate between the A and B stimuli. This finding is consistent with the idea that the perceptual space available to implant listeners is reduced compared with normal-hearing listeners. Compared with FDL, the degree of perceptual difference required to trigger automatic stream segregation (TCB) is generally much larger than that required simply to discriminate between stimuli (Moore and Gockel, 2002). Given this, a more limited perceptual space within which successive sounds can differ may reduce the ability of implant listeners to achieve stream segregation. It should be noted, however, that there are instances where the difference between segregation thresholds and discrimination thresholds can be smaller for normally-hearing listeners, particularly at high frequencies (e.g. see Rose and Moore, 2005).

Experiment 3 used an indirect measure of automatic stream segregation, i.e. a temporal discrimination task. Stream segregation should result in *worse* performance in this task despite the efforts of the listener, because it is hard to judge the relative timing of sounds that are heard in separate streams (e.g., Bregman and Campbell, 1971). Again, a clear effect of electrode separation was found, with greater separations leading to increased threshold delays. Stream segregation is cumulative (e.g. Bregman, 1978), and so if it had occurred a greater dependence on electrode separation should have been apparent for the longer sequences than for the short triplets. However, a similar dependence of threshold delays on electrode separation was seen for short duration ABA triplets as for longer sequences, with no interaction between electrode separation and sequence length. Once again, the results did not provide any clear evidence that automatic segregation had occurred. This finding contrasts with that reported by Hong and Turner (2006); differences in both the method of stimulus delivery and participant sampling may account for this discrepancy.

What then of cochlear implant listeners' abilities to attend selectively to parts of a stimulus sequence? As mentioned above, cochlear implant listeners do appear to show some ability to attend to target sounds and to ignore distracting ones in daily use of their devices. Experiment 4 evaluated whether this ability could be demonstrated in a melody discrimination task. Listeners discriminated between two simple melodies temporally interleaved with distractor tones on electrodes that were either selected from the same portion of the electrode array (i.e. from the same pitch range) or from a spatially separated part of the array (different pitch range). This was therefore a measure to which schema-based segregation, requiring active listening, would have contributed. On the basis of results from normally hearing listeners (e.g. Dowling, 1973), one might have expected that implant listeners could successfully perform this task if there were sufficient differences in the pitch ranges of the target melodies from the distracting tones, even when the target tones and distractors were of equal loudness. However, in practice these implant listeners were only able to achieve above-chance identification of simple target melodies when the distractors were substantially attenuated compared to the target sounds. This suggests that implant listeners are primarily reliant on *level* differences to select and attend to target sounds and to ignore distracting sounds; evidence that they were able to achieve this based on differences in pitch (or other perceptual differences produced by electrode separation, e.g. timbral brightness) was not found. Thus, efforts to improve signal-to-noise ratios in normal implant listening may be the most beneficial approach for helping implant users to cope in real-life noisy situations, for example using directional microphones, adaptive beam-forming, direct input from 'lapel' microphones, and FM systems, etc.

The extent to which cochlear implant listeners are capable of *simultaneous* grouping was explored in experiment 5. This was explored because the previous experiments had concerned only sequential grouping, and the ability to segregate concurrent sounds is clearly a critical element of everyday listening, for example when listening to speech in the presence of competing speech. Most potential cues for simultaneous grouping (e.g. harmonic relations, F0 differences) are not available at all via a cochlear implant, but abrupt changes in level are an exception. Normally hearing listeners are able to utilise a sudden change in a sound mixture as a cue to simultaneous grouping/segregation (Bregman et al., 1994a). This is important, because in real life abrupt changes in sound mixtures tend to signal the introduction of new sources, while gradual changes are more likely to be caused by changes in on-going sounds. Experiment 5 evaluated the ability of implant listeners to detect abrupt changes in the level of individual components in a complex tone, and to use those changes to perceptually isolate them from the rest of the complex. This was achieved by a temporary change in the stimulus level applied to one channel, and subsequently to another channel, when five implant channels were concurrently active. The results suggested that most implant listeners rely mainly on detecting peaks in energy as a cue for concurrent segregation, and do not benefit greatly from *abrupt* (as opposed to gradual) increases in level. The size of increments needed was substantially greater than in normal hearing, and implant listeners were mostly unable to use decrements in level (followed by a return to the previous level) as a segregation cue. One listener did perform on this task much more like a normally hearing listener, with an ability to detect very small increments. Her performance was so good that it was not possible to demonstrate any advantage for sudden-onset increments, but she did show an advantage for sudden *offsets* in decrements (i.e. rapid amplitude rises). This indicates that it is *possible* for an exceptional implant listener to give similar performance on such a task to that seen in normal hearing. Reasons for why her performance differed so strikingly from the others are difficult to evaluate, but it is interesting to note that she achieved 100% correct open-set speech recognition with her implant and so is clearly a 'star' implant user; also, she did have a relatively recent onset of profound deafness before implantation so that good neural survival in the implanted ear would be expected.

Limitations in the quality of spectral information that can be conveyed by cochlear implants, and the problem of channel interaction caused by spread of excitation from individual electrodes to their neighbours, have been cited frequently as major contributors to the limitations on their performance, especially in noise (e.g. Fu and Nogaki, 2004). As the ability to segregate sounds is also thought to be related to perceptual dissimilarities between them (e.g., Moore and Gockel, 2002), it is clearly important to understand and if possible evaluate the extent of the channel interaction problem. To this end, experiment 6 used an objective, neurophysiological measure of spread of excitation in the cochlea in implant users based on recording of the ECAP (Electrically Evoked Compound Action Potential) response using the NRT ™ system. This illustrated the often considerable overlap in excitation patterns between electrodes in the array – suggesting that the desired focused and discrete stimulation on individual electrodes with minimal overlap in areas of neural excitation was rarely apparent. Analysis of the ECAP data in relation to the results of experiment 1 suggested a significant relationship between the amplitude of ECAP responses (recorded using the 'Variable Masker, Fixed Recording' method) and reported segregation, i.e. a *greater* degree of overlapping stimulation between electrodes (resulting in larger ECAP responses) was associated with *less* reported segregation. This was encouraging, as it suggested a physiological basis for perceptual differences, which led to differences in responses on the segregation task. A somewhat weaker correlation was found with a measure of channel discrimination derived from electrode pitch rankings in experiment 2. However, a slightly stronger association of the ECAP data with the reported *confidence* in pitch ranking judgements was seen (i.e. larger ECAP responses, resulting from greater overlap of stimulation, were correlated with lower confidence in electrode pitch-ranking judgements). This is broadly consistent with the idea that *increased overlap* of excitation patterns resulting from stimulation of two different electrodes in sequence leads to *reduced* perceptual differences (i.e. based on pitch or timbre) between them.

### 9.3 Conclusions and future directions

A variety of experimental approaches have been used here to evaluate auditory grouping in cochlear implant listeners. No convincing evidence of *automatic* sequential segregation in these listeners was found on the tasks used, including both subjective and objective measures. It is not possible to conclude from this that implant listeners *never* experience primitive, automatic grouping of sounds in normal use of their devices, but these results suggest it may be a rare experience, as in normally hearing listeners performance on such tasks provides clear evidence of the operation of grouping processes. Furthermore, the cochlear implant listeners tested performed equally poorly on a task requiring selection of target sounds from distractors; they relied essentially on substantial differences in *level* between target and interfering sounds to perform schema-based segregation, and were apparently unable to achieve this using other perceptual differences, such as differences in pitch range. Similarly, changes in level were the primary cue used to detect tones 'popping out' from a complex tone in the simultaneous segregation task, with no clear evidence of any ability to use sudden onset as a segregation cue (other than for one exceptional listener).

Measures of spread of excitation in the cochlea based on ECAP recordings demonstrated the substantial overlap in neural activation patterns produced by stimulation of electrodes near to each other in the electrode array; this is consistent with the idea that spectral 'smearing' and channel interaction are major contributors to the limitations and variability of implant performance that are found. Also, the correlations between these objective measure of channel interaction and subjective reports of segregation or confidence in pitch ranking judgements suggests a physiological explanation for the results seen in those psychophysical tasks.

Some cues that are available to normally hearing listeners for automatic or schema-based grouping (e.g. differences in F0) are not expected to be available to implant listeners. Others, that *might* in principle be available to them, seem in practice to be of little use. Differences in pitch or timbre resulting from activation of different electrodes – even when they are widely separated along the array – do not seem to be an effective cue for either automatic or schema-based segregation. Although stimulation of different electrodes does produce perceptual differences, it appears that these are generally insufficient to facilitate the perceptual grouping of sounds. This is consistent with the idea that cochlear implant listeners have a reduced perceptual space within which sounds can differ from one another. In the absence of *automatic* stream segregation, they are reliant on schema-based selection of target sounds from the background, and even this may depend primarily on differences in level. In fact, many implant listeners appear to default to perceiving integrated percepts.

This merits a consideration of the implications for improvements to implant design. Much research effort is already aimed at improving selectivity of stimulation, as it is recognised that channel interaction caused by spread of excitation in the cochlea is a major limiting factor on performance. If more discrete stimulation of individual electrodes was achieved, this might allow better transmission of pitch or other perceptual differences between different electrodes (bearing in mind the limits on the number of effective channels that are available to cochlear implant listeners, as discussed in chapter 2). This should also increase the potential for either automatic or schema-based stream segregation based on pitch or timbre. Meanwhile, design of speech processors is frequently focused on improving the signal-to-noise ratio so that target speech has increased intelligibility in noise. Increasing the number of 'virtual channels' using current-steering to elicit a higher number of pitch percepts across the electrode array may sound like a promising way forward, but since it does not provide greater frequency selectivity it is unlikely to assist with the segregation of sounds or otherwise aid auditory scene analysis.

Research aimed at evaluation of auditory grouping has pitfalls; careful design of both tasks and stimuli is required in order to provide genuine and convincing evidence of stream segregation or other grouping processes. Results of previous studies with cochlear implant listeners have often been consistent with stream segregation, without necessarily providing unequivocal evidence that it has in fact occurred.

# References

1. Anstis,S. and Saida S (1985): Adaptation to auditory streaming of frequency-modulated tones. *Journal of Experimental Psychology, Human Perception and Performance*, 11:257-271.

2. Arehart,K.H., King,C.A., and McLean-Mudgett,K.S. (1997): Role of fundamental frequency differences in the perceptual separation of competing vowel sounds by listeners with normal hearing and listeners with hearing loss. *Journal of Speech, Language, and Hearing Research*, 40:1434-1444.

3. Assmann, P.F., Summerfield Q (1989): Modelling the perception of concurrent vowels: vowels with the same fundamental frequency. *Journal of the Acoustical Society of America, 85 (1): 327-338*

4. Baskent,D.B. (2006): Speech recognition in normal hearing and sensorineural hearing loss as a function of the number of spectral channels. *Journal of the Acoustical Society of America*, 120:2908-2925.

5. Beauvois,M.W. and Meddis R (1996): Computer simulation of auditory stream segregation in alternating-tone sequences. *Journal of the Acoustical Society of America*, 99:2270-2280.

6. Beauvois,M.W. and Meddis R (1997): Time decay of auditory stream biasing. *Perception and Psychophysics*, 59:81-86.

7. Bey,C. and McAdams,S. (2002): Schema-based processing in auditory scene analysis. *Perception and Psychophysics*, 64:844-854.

8. Bird,J. and Darwin,C.J. (1998): Effects of a difference in fundamental frequency in separating two sentences. In: Palmer, A.R., Rees, A., Summerfield, A.Q., Meddis, R. (Eds.), Psychophysical and Physiological Advances in Hearing. Whurr Publishers, London, 263-269.

9. Bregman,A.S. and Campbell,J. (1971): Primary auditory stream segregation and the perception of order in rapid sequences of tones. *Journal of Experimental Psychology*, 89:244-249.

10. Bregman,A.S. (1978): Auditory streaming is cumulative. *Journal of Experimental Psychology, Human Perception and Performance*, 4:380-387.

11. Bregman,A.S. (1990): Auditory Scene Analysis. *MIT Press, Cambridge, Mass., London.*

12. Bregman,A.S., Ahad,P., Kim,J., and Melnerich,L. (1994a): Resetting the pitch-analysis system: 1. Effects of rise times of tones in noise backgrounds or of harmonics in a complex tone. *Perception and Psychophysics*, 56(2):155-162.

13. Bregman,A.S. and Ahad,P. (1994b): Resetting the pitch-analysis system. 2. Role of sudden onsets and offsets in the perception of individual components in a cluster of overlapping tones. *Journal of the Acoustical Society of America*, 96(5):2694-2703.

14. Brokx,J.P. and Nooteboom,S.G. (1982): Intonation and the perceptual separation of simultaneous voices. *Journal of Phonetics*, 10:23-36.

15. Butler,G. and McManus,F.(2000): Psychology: A very short introduction. Oxford University Press, Oxford.

16. Carlyon,R.P., Plack,C.J., Fantini,D.A., and Cusack,R. (2003): Cross-modal and non sensory influences on auditory streaming. *Perception*, 32:1393-1402.

17. Carlyon,R.P., Long,C.J., Deeks,J., and McKay,C.M. (2007): Concurrent sound segregation in electric and acoustic hearing. *Journal of the Association for Research in Otolaryngology*, 8:119-133.

18. Carlyon,R.P., Cusack,R., Foxton,J.M., and Robertson,I.H. (2001): Effects of attention and unilateral neglect on auditory stream segregation. *Journal of Experimental Psychology, Human Perception and Performance*, 27:115-127.

19. Chatterjee M. and Galvin,J.J. (2002): Auditory streaming in cochlear implant listeners. *Journal of the Acoustical Society of America*, 111(5): p.2429

20. Chatterjee,M. and Shannon,R.V. (1998): Forward masked excitation patterns in multielectrode electrical stimulation. *Journal of the Acoustical Society of America*, 103(5):2565-2572.

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

22. Cherry,E.C. (1953): Some experiments on the recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America*, 25:975-979.

23. Chung K., Killion, M.C., Christensen, L.A. (2002): Ranking hearing aid input-output functions for understanding low-, conversational-, and high-level speech in multitalker babble. *Journal of Speech, Language and Hearing Research, 50(2): 304-322*

24. Cohen,L.T., Richardson,L.M., Saunders,E., and Cowan,R.S.C. (2003): Spatial spread of neural excitation in cochlear implant recipients: comparison of improved ECAP method and psychophysical forward masking. *Hearing Research*, 179:72-87.

25. Cullington H.E. and Zeng,F.-G. (2008): Speech recognition with varying numbers and types of competing talkers by normal-hearing, cochlear-implant, and implant simulation subjects. *Journal of the Acoustical Society of America*, 123:450-461.

26. Cusack,R. and Roberts,B. (1999): Effects of similarity in bandwidth on the auditory sequential streaming of two-tone complexes. *Perception*, 28:1281-1289.

27. Cusack,R. and Roberts,B. (2000): Effects of differences in timbre on sequential grouping. *Perception and Psychophysics*, 62:1112-1120.

28. Cusack,R. and Roberts,B. (2004): Effects of differences in the pattern of amplitude envelopes across harmonics on auditory stream segregation. *Hearing Research*, 193:95-104.

29. Cusack,R., Deeks,J., Aikman,G., and Carlyon,R.P. (2004): Effects of location,frequency region,and time course of selective attention on auditory scene analysis. *Journal of Experimental Psychology, Human Perception and Performance*, 30(4):643-656.

30. Cusack,R. and Carlyon,R.P. (2004): Auditory perceptual organisation inside and outside the laboratory. *Chapter in: Ecological Psychoacoustics*, 15-45.

31. Dannenbring,G.L. and Bregman,A.S. (1976): Stream segregation and the illusion of overlap. *Journal of Experimental Psychology, Human Perception and Performance*, 2:544-555.

32. Darwin C.J. and Ciocca,V. (1992): Grouping in pitch perception: effects of onset asynchrony and ear of presentation of a mistuned component. *Journal of the Acoustical Society of America*, 91:3381-3390.

33. Darwin,C.J. (1984): Perceiving vowels in the presence of another sound: constraints on formant perception. *Journal of the Acoustical Society of America*, 76:1636-1647.

34. Darwin,C.J. (2005): Pitch and Auditory grouping. *In: Pitch, Neural coding and perception, Plack,C.J., Oxenham,A.J., Fay,R.R. and Popper A.N. Springer*, 279-305.

35. David,E.E., Ostroff,J.M., Shipp,D., Nedzelski,J.M., Chen,J.M., Parnes,L.S., Zimmerman,K., Schramm,D., and Seguin,C. (2003): Speech coding strategies and revised cochlear implant

candidacy: an analysis of post-implant performance. *Otology and Neurotology*, 24:228-233.

36. Deeks,J. and Carlyon,R.P. (2004): Simulations of cochlear implant hearing using filtered harmonic complexes: Implications for concurrent sound segregation. *Journal of the Acoustical Society of America*, 115(4):1736-1746.

37. Donaldson,G.S. and Nelson,D.A. (2000): Place-pitch sensitivity and its relation to consonant recognition by cochlear implant listeners using the MPEAK and SPEAK speech processing strategies. *Journal of the Acoustical Society of America*, 107(3):1645-1658.

38. Dorman,M.F., Loizou,P.C., Fitzke,J., and Tu,Z. (1998): The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear-implant signal processors with 6-20 channels. *Journal of the Acoustical Society of America*, 104(6):3583-3585.

39. Dowling,W.J. (1973): The perception of interleaved melodies. *Cognitive Psychology*, 5:322-337.

40. Friesen,L.M., Shannon,R.V., Baskent,D.B., and Wang,X. (2001): Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants. *Journal of the Acoustical Society of America*, 110(2):1150-1163.

41. Fu,Q.-J. and Nogaki,G. (2004): Noise susceptibility of cochlear implant users: the role of spectral resolution and smearing. *Journal of the Association for Research in Otolaryngology*, 6:19-27.

42. Gaudrain,E., Grimault N., Healy,E.W., and Bera,J.-C. (2007): Effect of spectral smearing on the perceptual segregation of vowel sequences. *Hearing Research*, 231:32-41.

43. Glasberg,B.R. and Moore B.C.J. (1990): Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, 47:103-138.

44. Greenwood D.D. (1990): A cochlear frequency-position function for several species- 29 years later. *Journal of the Acoustical Society of America*, 87:2592-2605.

45. Grimault,N., Micheyl,C., Carlyon,R.P., Arthaud,P., and Collet,L. (2001): Perceptual auditory stream segregation of sequences of complex sounds in subjects with normal and impaired hearing. *British Journal of Audiology*, 35:173-182.

46. Grose,J.H., Hall,J.W., Buss,E., and Hatch,D. (2001): Gap detection for similar and dissimilar gap markers. *Journal of the Acoustical Society of America*, 109(4):1587-1595.

47. Hanekom,J.J. and Shannon,R.V. (1998): Gap detection as a measure of electrode interaction in cochlear implants. *Journal of the Acoustical Society of America*, 104(4):2372-2384.

48. Hartmann,W.M., McAdams,S., and Smith,B.K. (1990): Hearing a mistuned harmonic in an otherwise periodic complex tone. *Journal of the Acoustical Society of America*, 88:1712-1724.

49. Hartmann,W.M. and Johnson,D. (1991): Stream segregation and peripheral channeling. *Music perception*, 9:155-184.

50. Hawley,M.L., Litovsky,R.Y., and Culling,J.F. (2004): The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer. *Journal of the Acoustical Society of America*, 115(2):833-843.

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

52. Hughes,M.L. and Abbas P.J. (2006): The relation between electrophysiologic channel interaction and electrode pitch ranking in cochlear implant listeners. *Journal of the Acoustical Society of America*, 119:1527-1537.

53. Keppel,G. and Wickens, T.D. (2004) Design and Analysis, a Researcher's Handbook, 4th edition, Pearson Prentice Hall, New Jersey

54. Kidd,G., Arbogast,T.L., Mason,C.R., and Gallun,F.J. (2005): The advantage of knowing where to listen. *Journal of the Acoustical Society of America*, 118:3804-3815.

55. Koch,D.B., Downing,M., Osberger M.J., and Litvak,L. (2007): Using current steering to increase spectral resolution in CII and HiRes 90K users. *Ear and Hearing*, 28:38-41.

56. Kong,Y.-Y., Cruz,R., Ackland Jones,J., and Zeng,F.-G. (2004): Music perception with temporal cues in acoustic and electric hearing. *Ear and Hearing*, 25(2):173-185.

57. Kubovy,M. and Daniel,J.E. (1983): Pitch segregation by interaural phase, by momentary amplitude disparity, and by monaural phase. *Journal of the Audio Engineering Society*, 31:630-634.

58. Lacherez,P., Seah,E.L., and Sanderson,P. (2007): Overlapping melodic alarms are almost indiscriminable. *Human Factors*, 49:637-645.

59. Laneau,J., Wouters,J., and Moonen,M. (2006): Improved music perception with explicit pitch coding in cochlear implants. *Audiology and Neurootolology.*, 11:38-52.

60. Litvak,L.M.. Spahr, A.J., Saoji, A.A., and Fridman, G.Y (2007a): Relationship between perception of spectral ripple and speech recognition in cochlear implant and vocoder listeners. *Journal of the Acoustical Society of America*, 122 (2), 982-991

61. Litvak, L.M., Spahr, A.J., Emadi, G. (2007b): Loudness growth observed under partially tripolar stimulation: model and data from cochlear implant listeners. *Journal of the Acoustical Society of America*, 122 (2), 967-981.

62. Luts, H., Maj, J.B., Soede,W., Wouters, J (2004): Better speech perception in noise with an assistive multimicrophone array for hearing aids. *Ear and Hearing, 25(5), 411-420*

63. Massaro,D.W. (1977): Perceptual grouping in audition. *Perception*, 6:541-553.

64. McDermott,H.J. (2004): Music perception with cochlear implants: a review. *Trends in Amplification*, 8(2):49-82.

65. Miller,G.A. and Heise,G.A. (1950): The trill threshold. *Journal of the Acoustical Society of America*, 22:637-638.

66. Moore, B.C.J. (Ed) (1995): Hearing. *Handbook of Perception and Cognition, 2nd edition. Series editors Carterette, E.C and Friedman, M.P., Academic Press*

67. Moore B.C.J., Glasberg,B.R., and Peters,R.W. (1986): Thresholds for hearing mistuned partials as separate tones in harmonic complexes. *Journal of the Acoustical Society of America*, 80:479-483.

68. Moore,B.C.J. and Gockel,H. (2002): Factors influencing sequential stream segregation. *Acta Acustica*, 88:320-332.

69. Moore,B.C.J. and Carlyon,R.P. (2005): Perception of pitch by people with cochlear hearing loss and by cochlear implant users. *Springer handbook of auditory research, Vol. 24: Pitch, Neural coding and perception,* Plack, C.J., Oxenham, A.J., Fay, R.R., Popper, A.N. (Eds.)

70. Nascimento,L.T. and Bevilacqua,M.C. (2005): Evaluation of speech perception in noise in cochlear implanted adults. *Revista brasileira de otorrinolaringologia* 71:432-437.

71. Neff,D.L., Jesteadt,W., and Brown,E. (1982): The relation between gap discrimination and auditory stream segregation. *Perception and Psychophysics*, 31:493-501.

72. Nelson,D.A., Van Tassell,D.J., Schroder,A.C., Soli,S., and Levine,S. (1995): Electrode ranking of 'place pitch' and speech recognition in electrical hearing. *Journal of the Acoustical Society of America*, 98(4):1987-1999.

73. Nelson,P.B. and Jin,S.-H. (2002): Understanding speech in single-talker interference: Normal-hearing listeners and cochlear implant users. *Journal of the Acoustical Society of America*, 111 (5):p.2429

74. Nelson,P.B., Jin,S.-H., Carney,A.E., and Nelson,D. (2003): Understanding speech in modulated interference: Cochlear implant users and normal-hearing listeners. *Journal of the Acoustical Society of America*, 113(2):961-968.

75. Roberts,B. and Moore,B.C.J. (1991): The influence of extraneous sounds on the perceptual estimation of first-formant frequency in vowels under conditions of asynchrony. *Journal of the Acoustical Society of America*, 89:2922-2932.

76. Roberts,B., Glasberg,B.R., and Moore,B.C.J. (2002): Primitive stream segregation of tone sequences without differences in fundamental frequency or passband. *Journal of the Acoustical Society of America*, 112:2074-2085.

77. Roberts,B., Glasberg,B.R., and Moore,B.C.J. (2008): Effects of the build-up and resetting of auditory stream segregation on temporal discrimination. *Journal of Experimental Psychology, Human Perception and Performance*, In press

78. Rogers,W.L. and Bregman,A.S. (1993): An experimental evaluation of three theories of auditory stream segregation. *Perception and Psychophysics*, 53:179-189.

79. Rogers,W.L. and Bregman,R.S. (1998): Cumulation of the tendency to segregate auditory streams: Resetting by changes in location and loudness. *Perception and Psychophysics*, 60:1216-1227.

80. Rose,M.M. and Moore,B.C.J. (1997): Perceptual grouping of tone sequences by normally hearing and hearing impaired listeners. *Journal of the Acoustical Society of America*, 102:1768-1778.

81. Rose,M.M. and Moore,B.C.J. (2005): The relationship between stream segregation and frequency and frequency discrimination in normally hearing and hearing-impaired subjects. *Hearing Research*, 204:16-28.

82. Scheffers,M.T.M. (1982): The role of pitch in the perceptual separation of simultaneous vowels II. *IPO Annual Progress Report*, 17:41-45.

83. Shannon,R.V., Zeng,F.-G., Kamath,V., Wygonski,J., and Ekelid,M. (1995): Speech recognition with primarily temporal cues. *Science*, 270:303-304.

84. Shannon,R.V., Fu,Q.-J., and Galvin,J.J. (2004): The number of spectral channels required for speech recognition depends on the difficulty of the listening situation. *Acta Otolaryngologica*, Supp.552:1-5.

85. Shannon,R.V. (2005): Speech and music have different requirements for spectral resolution. *International Review of Neurobiology*, 70:121-134.

86. Singh,P.G. and Bregman,A.S. (1997): The influence of different timbre attributes on the perceptual segregation of complex-tone sequences. *Journal of the Acoustical Society of America*, 102:1943-1952.

87. Spriet,A., Van Deun,L., Eftaxiadis,K., Laneau,J., Moonen,M., van Dijk,B., van Wieringen,A., and Wouters,J. (2007): Speech understanding in background noise with the two-microphone adaptive beamformer BEAM in the Nucleus Freedom Cochlear Implant System. *Ear and Hearing*, 28(1):62-72.

88. Stainsby,T.H., Moore,B.C.J., and Glasberg,B.R. (2004): Auditory streaming based on temporal structure in hearing-impaired listeners. *Hearing Research*, 192:119-130.

89. Stickney,G.S., Zeng,F.-G., Litovsky,R.Y., and Assmann,P.F. (2004): Cochlear implant speech recognition with speech maskers. *Journal of the Acoustical Society of America*, 116:1081-1091.

90. Stickney,G.S., Loizou,P.C., Mishra,L.N., Assmann,P.F., Shannon,R.V., and Opie,J. (2006): Effects of electrode design and configuration on channel interactions. *Hearing Research*, 211:33-45.

91. Stickney,G.S., Assmann,P.F., Chang J., and Zeng,F.-G. (2007): Effects of cochlear implant processing and fundamental frequency on the intelligibility of competing sentences. *Journal of the Acoustical Society of America*, 122:1069-1078.

92. van Noorden,L.P.A.S. (1975): Temporal coherence in the perception of tone sequences. *Unpublished PhD thesis, Eindhoven University of Technology*

93. Vliegen,J. and Oxenham,A.J. (1999): Sequential stream segregation in the absence of spectral cues. *Journal of the Acoustical Society of America*, 105:339-346.

94. Vliegen,J., Moore,B.C.J., and Oxenham,A.J. (1999): The role of spectral and periodicity cues in auditory stream segregation,

measured using a temporal discrimination task. *Journal of the Acoustical Society of America*, 106:938-945.

95. Warren,R.M., Obusek,C.J., Farmer,R.M., and Warren,R.P. (1969): Auditory sequence: Confusion of patterns other than speech or music. *Science*, 164:586-587.

96. Wilson,B.S. (2006): Speech processing strategies. *Chapter in: Cochlear Implants- a Practical Guide, Cooper H.R. and Craddock L.C (eds)*, 21-69. Whurr, London

97. Wilson,B.S. and Dorman,M.F. (2007): The surprising performance of present-day cochlear implants. *IEEE Transactions on Biomedical Engineering, 54 (6 part 1):*969-972.

## Previous presentations of data from this thesis

1. Cooper H.R., and Roberts, B. (2003) Auditory sequential grouping in cochlear implant listeners: Preliminary results.
Poster presented at the British Society of Audiology short papers meeting on experimental studies of hearing and deafness, University of Nottingham, September

2. Cooper H.R., and Roberts, B. (2004) Auditory Stream Segregation and Channel Discrimination in Cochlear Implant Listeners.
Poster presented at the British Society of Audiology short papers meeting on experimental studies of hearing and deafness, University College London, September

3. Cooper, H.R., and Roberts, B. (2005). "Auditory stream segregation and channel discrimination in cochlear implant listeners." Conference on Implantable Auditory Prostheses, 30 July - 4 August, Asilomar, Pacific Grove, California, p.119.

4. Cooper H.R., and Roberts, B. (2006) Auditory stream segregation in cochlear implant listeners. Paper presented at the British Academy of Audiology annual conference, Telford, November

5. Cooper H.R., and Roberts, B. (2007) Auditory stream segregation of tone sequences in cochlear implant listeners. *Hearing Research*, 225, 11-24

6. Cooper, H.R. and Roberts, B. (2007) Interleaved melody recognition in cochlear implant listeners. Poster presented at the British Cochlear Implant Group meeting in Ireland, Dublin, April

7. Cooper, H.R. and Roberts, B. (2007) Can cochlear implant listeners use abrupt changes in level to pick out components from a complex tone?
Poster presented at the British Society of Audiology short papers meeting on experimental studies of hearing and deafness, University College London, September

# APPENDICES

## APPENDIX 1:

## EXPERIMENT 1: INSTRUCTIONS

The study which you are contributing to is concerned with the ability of people with cochlear implants to hear different sounds through their implant as either one sound or separated out into two distinct sounds. The results of these experiments will help us to understand more about how the hearing system in our brain works, and how this may be the same or different in cochlear implant users compared with people with normal hearing.

The experiment involves listening to a series of tones, which are generated by computer and fed into the implant. The tones will be of different pitches. Sometimes you will hear a distinct 'galloping' rhythm and it is impossible to hear the two alternating tones as two distinct sounds. At other times, although the tones may start off sounding like one sound with the 'galloping' rhythm, they will appear to separate out into two distinct 'streams' and you can listen to either one or other of the tones of differing pitch.

The screen will always say "1 sound-gallop" at the start. If you can't hear two separate sounds or pitches, and there is a galloping rhythm, don't touch the space bar. If you can hear two separate sounds, and there is no galloping rhythm, hit the space bar and the screen will change to say "2 separate sounds".

When you can't hear two separate sounds or pitches, and there is a galloping rhythm, make sure the screen says "1 sound-gallop".NB: the space bar does not change the sounds you hear; it only changes what is displayed on the screen.

Sometimes it will be clear which of the two choices you are hearing, but sometimes what you hear may be ambiguous, and may seem to 'flip' between the '1 sound-gallop' and the '2 separate sounds'. All you need to do is hit the space bar each time the sounds you hear *change*, so that the screen always shows what you are hearing at that moment. You can hit the space bar as many times as you wish. Every time you touch the space bar, the display on the screen changes from one message to the other.

NB: There is no 'right' or 'wrong' response in this experiment- the idea is to record what *you* are hearing.

Practice will be given before the actual test begins. I hope that you find the experiment interesting, and thankyou very much for helping with this research.

**Please ask if you have any questions**

## APPENDIX 2

## EXPERIMENT 2: INSTRUCTIONS

The purpose of this experiment is to explore how you perceive different sounds as different pitches through your implant.

On each trial, you will hear two sounds one after the other, with only a short pause in between each. Please listen to both, and then indicate which of the sounds (the first one or the second one) sounded higher pitched, brighter or sharper. Press key 1 if the first sound was higher and key 2 if th second sounds was higher.

Sometimes there may not be a clear difference in pitch, but one sound may sound 'brighter' or 'sharper'- this is a valid response, even if what you hear is not a clear pitch difference. Also, sometimes it may be very difficult to hear any difference at all between the two sounds; if this happens, please have a guess anyway. The experiment cannot continue until you respond.

After you have pressed key 1 or 2, you will be asked on the computer screen to state how confident you were about your choice, on a scale from 1 to 5. On this scale, a response of 1 implies that you were more or less guessing and were very unsure about your choice. A response of 5 means that you were absolutely sure about your choice. You can respond with any of the numbers between 1 and 5 along these lines:

1: Very unsure, really guessing
2: A little bit confident
3: Moderately confident
4: Very confident but not certain
5:Absolutely confident

Press one of the keys from 1 to 5 to indicate your response. No other keys will do anything. The experiment will not proceed until you have responded both the first time (1 or 2) and the second time (1,2,3,4 or 5).

There are 44 pairs of sounds in each block. We will need to complete 10 blocks in total.

**THANKYOU FOR HELPING WITH THIS STUDY**

**Please ask if you have any questions**

## APPENDIX 3

## EXPERIMENT 3 (TASK 1): INSTRUCTIONS

In this task, on each trial you will hear two short sequences of sounds, one after the other. Each sequence lasts about 3 seconds, and is made up of a series of tones of various pitches. In one of the sequences, the rhythm of the tones is entirely regular throughout the sequence. In the other, the rhythm of the tones starts off regular but then becomes irregular towards the end of the sequence. Your task is to listen to both sequences, then decide which one is **irregular.** In every case, one of the sequences is regular and the other is irregular- but the order in which they are played varies at random.

Sometimes, it is quite easy to say which of the sequences is irregular- but sometimes they may sound very similar, so it is very difficult to say which one is which. However, you will need to choose one or the other (1 or 2) every time, even if you are not really sure or are just guessing.

After you have responded by pressing either the 1 or 2 key on the computer keyboard, you will be given feedback on the screen in front of you, stating whether your response was correct or incorrect. Don't worry if you get some of them incorrect- this is completely normal, and an analysis of which ones you get correct and which incorrect provides valuable information.

There are 120 trials in each block, and each block takes about 17 minutes to complete. We need to complete 10 blocks altogether in order to collect enough data to analyse.

You will be given sufficient practice before starting the test, and can stop for a break at any time in between blocks.

Instructions for Task 2: as above, but each sequence lasts about 1 second and consists of only three brief sounds or 'triplets'. In a *regular* version of a triplet, the second sound occurs at the mid-point in time between the first and third sounds. In an *irregular* version, the middle sound is slightly delayed so that the triplet sounds irregular. Please indicate as before which triplet, first (1) or second (2) sounds *irregular.*

**THANKYOU FOR HELPING WITH THIS STUDY**
**Please ask if you have any questions**

## APPENDIX 4

## EXPERIMENT 4 :INSTRUCTIONS

In this experiment, you will be asked to listen to a short sequence of tones which form a simple melody. In each case the sequence you will be presented with will be either sequence 1 or sequence 2. You will be given examples and training with both sequences so that you are familiar with them before the experiment begins. All sounds will be given to you at the same comfortable listening level.

Your task is to listen to the tone sequence and then indicate whether you heard sequence 1 or sequence 2 by pressing either the 1 or 2 key on the keyboard. You will have time to think, and the next tone sequence will not be presented to you until you have responded to the previous one.

In the second part of the experiment, the tone sequence you are listening for will be mixed up with some other 'distractor' sounds which may make the task more difficult. Your task is if possible to ignore the interfering sounds and to listen for either sequence 1 or sequence 2 as before and respond accordingly.

**THANKYOU FOR HELPING WITH THIS STUDY**
**Please ask if you have any questions**

## APPENDIX 5

## EXPERIMENT 5 :INSTRUCTIONS

In this experiment, you will be asked to listen to a short sound (3.5 seconds long). This may sound predominantly like noise or a buzz, but you should also be able to hear two brief additional sounds, one after the other, during the overall sound. One of these, either the first or the second, should have a higher pitch than the other, and your task is to report whether the resulting pitch pattern is falling (i.e. high then lower pitch) or rising (i.e. low then higher pitch).

On the response box in front of you is a red button marked S. When you are ready, press this button in order to hear the next sound. After hearing the sound, please respond by pressing either button 1 if you heard a falling pitch pattern (i.e. high then lower pitch) or button 2 if you heard a rising pitch pattern (i.e. low then higher pitch). After responding, the button marked S will be re-lit for you to press as soon as you are ready to listen to the next sound. You will have time to think, and the next sound will not be presented to you until you have responded to the previous one and pressed the S button.

**THANKYOU FOR HELPING WITH THIS STUDY**
**Please ask if you have any questions**