

Auditory stream segregation in cochlear implant listeners: Measures based on temporal discrimination and interleaved melody recognition

Huw R. Cooper and Brian Roberts^{a)}

Psychology, School of Life and Health Sciences, Aston University, Birmingham B4 7ET, United Kingdom

(Received 7 April 2009; revised 7 July 2009; accepted 8 July 2009)

The evidence that cochlear implant listeners routinely experience stream segregation is limited and equivocal. Streaming in these listeners was explored using tone sequences matched to the center frequencies of the implant's 22 electrodes. Experiment 1 measured temporal discrimination for short (ABA triplet) and longer (12 AB cycles) sequences (tone/silence durations = 60/40 ms). Tone A stimulated electrode 11; tone B stimulated one of 14 electrodes. On each trial, one sequence remained isochronous, and tone B was delayed in the other; listeners had to identify the anisochronous interval. The delay was introduced in the second half of the longer sequences. Prior build-up of streaming should cause thresholds to rise more steeply with increasing electrode separation, but no interaction with sequence length was found. Experiment 2 required listeners to identify which of two target sequences was present when interleaved with distractors (tone/silence durations = 120/80 ms). Accuracy was high for isolated targets, but most listeners performed near chance when loudness-matched distractors were added, even when remote from the target. Only a substantial reduction in distractor level improved performance, and this effect did not interact with target-distractor separation. These results indicate that implantees often do not achieve stream segregation, even in relatively unchallenging tasks.

© 2009 Acoustical Society of America. [DOI: 10.1121/1.3203210]

PACS number(s): 43.66.Mk, 43.66.Ts, 43.66.Fe, 43.66.Hg [BCM]

Pages: 1975–1987

I. INTRODUCTION

Grouping together only those acoustic elements that arise from a common source is an important function of the auditory system, and the extent to which this is possible has a major impact on one's ability to recognize auditory objects in complex listening environments. Normal-hearing (NH) and cochlear implant (CI) listeners alike must solve this scene analysis problem (Bregman, 1990), but CI listeners are much less well equipped for this challenge owing to the impoverished sensory information available via their implants (e.g., Throckmorton and Collins, 2002; Fu and Nogaki, 2005). However, relatively few studies have investigated auditory grouping in CI listeners, and even fewer have used accuracy of performance as a measure of perceptual organization with implant listening. The experiments reported here used temporal discrimination and interleaved melody recognition to measure the extent to which CI listeners typically experience auditory stream segregation.

Bregman (1990) proposed two distinct processes in the construction of auditory representations: "primitive" scene analysis and schema-based selection. The first is driven by the incoming acoustic data and is often assumed to be involuntary and pre-attentive. The latter is assumed to be voluntary and to involve, through selective attention, 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). The distinction between these processes is reflected in the findings of van Noorden (1975). He examined the perceptual organization of sequences comprising alternating high- and low-frequency pure tones that differed in frequency separation and presentation rate (tone onset-to-onset time). van Noorden (1975) distinguished between the temporal coherence boundary (TCB) and the fission boundary (FB) by varying the instructions given to his participants, and hence their listening "set." He found that the sequence could not be heard as a single stream above the TCB presumably because of the obligatory operation of primitive stream segregation, whereas it could not be heard as two streams below the FB presumably because of the limits of attentional selection. The TCB was influenced by frequency separation and rate; larger separations were required to induce streaming at slower rates. In contrast, the FB showed little dependence on rate; sounds could be selected voluntarily from a larger set if their frequency separation was larger than a minimum critical value (a few semitones).

A. Temporal discrimination as a streaming measure in NH listeners

The perceptual properties of sound events are computed within streams but not across them. Therefore, one approach to measuring auditory stream segregation is to use a task for which performance will be affected adversely if streaming occurs. For example, performance is impaired in tasks that require listeners to count or report the order of sequences of pure tones if these sounds are heard to split into separate

^{a)}Author to whom correspondence should be addressed. Electronic mail: b.roberts@aston.ac.uk

streams on the basis of frequency separation or rate (Warren *et al.*, 1969; Bregman and Campbell, 1971). Similarly, stream segregation increases thresholds for detecting changes in rhythm arising from temporal asymmetries between sequentially presented tones (e.g., van Noorden, 1975; Vliegen *et al.*, 1999; Cusack and Roberts, 2000; Roberts *et al.*, 2002). Given that listeners must attempt to hear the sequence as a single stream to perform well in these tasks, the results should indicate the limits of their ability to overcome primitive stream segregation.

One important caveat regarding this approach merits note. Even for two isolated pure tones, for which one would expect little or no build-up in the tendency for stream segregation (Bregman, 1978), thresholds for detecting a temporal gap between them rise for NH listeners as the frequency separation increases (e.g., Grose *et al.*, 2001). Thus, the effects of frequency differences between tones on performance in gap discrimination tasks do not always reflect increased stream segregation. Indeed, gap detection thresholds have been shown to rise in CI listeners as the physical separation of stimulated electrode pairs is increased; Hanekom and Shannon (1998) used this as a measure of electrode interaction, not of streaming.¹ Therefore, an increase in temporal-discrimination thresholds with greater frequency separation between two stimuli cannot in itself be taken as firm evidence of stream segregation.

To demonstrate convincingly that changes in threshold reflect changes in stream segregation, it is necessary to observe a greater dependence of temporal judgments on frequency separation in situations where stream segregation is expected to be greater, i.e., when there has been sufficient time for the tendency for segregation to build up. This was first demonstrated by van Noorden (1975), who measured the just perceptible displacement in time (ΔT) of the B tones from the midpoint between neighboring A tones in long sequences of alternating AB tones and in short ABA “triplets.” There was a very clear dependence of ΔT on the frequency ratio between tones A and B for the long sequences, but there was far less dependence of ΔT on the AB ratio for the triplets.

B. Temporal discrimination as a streaming measure in implant listeners

To date, the only published example of this approach being applied to CI listeners is the study by Hong and Turner (2006). They used sequences of pure tones configured in the same way as those described by Roberts *et al.* (2002), where the target stimulus began with an isochronous portion to allow significant build-up of the strength of stream segregation before the anisochrony was introduced. Hence, the ability to detect the irregular rhythm was taken to indicate an absence of streaming. Stimuli were presented via a loudspeaker at 90 or 95 dB SPL (sound pressure level) to eight CI listeners, who used their normal speech processors and “maps” (speech processing strategies and stimulus levels), and were allowed to adjust their own volume or sensitivity settings if desired. Tone A was set at 200, 800, or 2000 Hz, while the frequency of tone B was set at a ratio $[(B-A)/A]$ of 0 (i.e.,

$A=B$), 0.01, 0.1, 0.5, 1.0, or 3.0. Using an adaptive two-interval two-alternative forced-choice task, they measured the threshold delay for tone B over this range of separations from each “base frequency” for tone A. The results were normalized relative to the threshold delay for the case where the ratio was 0. Hong and Turner (2006) found considerable variability in the slopes of the function relating threshold to frequency separation for their CI listeners. The AB frequency differences were converted into electrode separations based on the frequency allocations to each electrode for the individual CI listeners, and significant linear regressions were found between the normalized 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 three NH and three CI listeners at two ratios (0 and 0.5) using isolated ABA tone triplets (“short rhythm task”) as well as the longer sequences (“streaming task”). For the three base frequencies, they found that the normalized thresholds for the ratio of 0.5 were usually greater for the longer sequences than for the triplets. In effect, they found an interaction between tone sequence length and frequency separation, which they concluded was evidence that the task was a genuine measure of stream segregation. They went on to report correlations between normalized thresholds in their streaming task and speech reception thresholds in noise.

In contrast with Hong and Turner (2006), Cooper and Roberts (2007) argued that most CI listeners typically experience little or no auditory stream segregation. Listeners in their task were required to report whether they heard one or two streams when listening to sequences of alternating high and low tones. A significant relationship was found between electrode separation and reported segregation, but sequence rate had essentially no effect on listeners’ judgments. In contrast, rate changes usually have a major effect on streaming judgments by NH listeners (van Noorden, 1975). Furthermore, there was little evidence of the perceptual instability for intermediate electrode separations that would have been predicted if stream segregation were occurring (Anstis and Saida, 1985). A similar effect of electrode separation on reported segregation was observed by Chatterjee *et al.* (2006), but they did not explore the effect of changes in rate. Cooper and Roberts (2007) concluded that their results probably reflected simple channel discrimination rather than streaming judgments.

What might account for the discrepancy between the findings of these two studies? One possibility concerns the listening set of the participants in the study of Cooper and Roberts (2007), for which listeners were not instructed to try and hear the stimuli in any particular way. However, as noted by Cooper (2008), while it is true that the dependence of reported segregation on sequence rate may not be as strong under neutral instructions as when listeners are asked to try and hear a single stream, there is clear experimental evidence of this dependence under neutral instructions in NH listeners (e.g., Anstis and Saida, 1985, experiment 1). Another possibility concerns underlying differences between measures of

streaming based on subjective report and on temporal discrimination. Although these measures are often regarded as closely related (e.g., [Roberts et al., 2002](#)), there is recent evidence of discrepancies between them [compare the results of [Roberts et al. \(2008\)](#) with those of [Rogers and Bregman \(1993, 1998\)](#)]. However, so far these discrepancies appear to be restricted to studies concerned specifically with the resetting of the build-up of stream segregation, which does not apply here.

The results reported by [Hong and Turner \(2006\)](#) should be treated with some caution because there are a number of caveats relating to their experimental methods. First, stimuli were presented via loudspeaker rather than by direct stimulation, and so the effects of each listener's speech processing strategy on the resulting patterns of electrical stimulation are not entirely clear. Indeed, unlike our listeners, a mixture of different implant systems and speech-coding strategies was used among their participants, 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. Second, frequency separation was the experimental parameter directly manipulated, and this was only converted into electrode separation afterward. Again, this inevitably leads to some uncertainty about the precise stimuli that were delivered to the electrodes of the listeners' implants. Finally, only three CI listeners (one user of the 22-channel Nucleus implant and two users of the 16-channel Clarion implant) participated in their critical second experiment, which compared longer sequences and triplets, and only two frequency separations were tested (ratios of 0 and 0.5). As noted by [Cooper and Roberts \(2007\)](#), results based on more precisely controlled stimuli for a wider range of electrode separations, and from more participants, would be required to provide convincing evidence of involuntary stream segregation in the majority of implant listeners. Therefore, experiment 1 reported here revisited the use of a temporal-discrimination task to explore stream segregation in CI listeners. The experiment was designed to avoid some of the limitations associated with [Hong and Turner's \(2006\)](#) methods.

C. Interleaved melody recognition and schema-based stream segregation

One example of a situation that would be expected to be influenced by schema-based streaming is when the listener has to select and recognize a familiar melody or pattern of tones from a sequence containing interleaved interfering sounds (distractors). [Dowling \(1973\)](#) described a task of this type, in which he played to NH listeners pairs of familiar melodies (e.g., "Happy Birthday") that were temporally interleaved, i.e., where the tones of one melody alternated with the tones of another. The listeners' task was to name either of the melodies. He found that their pitch ranges should not overlap for identification to be successful.

[Cusack and Roberts \(2000\)](#) modified [Dowling's \(1973\)](#) task such that an arbitrary six-tone melody was heard once in isolation, followed by two intervals in which the melody was interleaved with distractors. The melody was identical to the isolated version in one interval and was modified in the other

interval; the task was to identify the interval containing the exact match. Differences between targets and distractors in pitch range and in timbre (pure tones vs narrow-band noises) both improved performance considerably. [Bey and McAdams \(2002, 2003\)](#) have since used a similar/different task to compare the case where an isolated arbitrary melody (comparison) is heard before or after the interleaved stimulus. Listeners were better able to judge whether the interleaved melody was identical to the comparison melody when the comparison was presented beforehand, indicating that prior knowledge of the melody to be extracted from the mixture enabled a contribution from a schema-based process for segregation. Thus, tasks requiring the recognition of interleaved melodies have been effective in experimental studies of auditory stream segregation in NH listeners. Such tasks are likely to involve both primitive and schema-driven segregation processes but are probably dominated by the latter.

We are aware of only one study that has explored the ability of CI listeners to select a subset of acoustic elements from a longer sequence, and only a short summary of this study has been published ([Chatterjee and Galvin, 2002](#)). These authors used repeating patterns of loudness-matched stimuli that were composed of two or three different tones (tone duration = 50 ms; inter-tone interval = 50 ms), and the tonotopic distance (electrode pair separation) between these tones was varied. Before each test sequence, listeners heard a "preview" sequence with a rhythm corresponding to a subset of the tones and were asked to judge whether or not they could hear this rhythm within the test sequence. The proportion of positive responses was taken as evidence that the subset of elements could be heard out as a separate perceptual stream. [Chatterjee and Galvin \(2002\)](#) found that this proportion was very dependent on tonotopic distance, as would be expected for NH listeners. Although their method was described as a yes/no task, the preview was always contained within the longer sequence, and so the task was subjective in nature. Experiment 2 reported here used an interleaved melody task to evaluate schema-based segregation in CI listeners using an accuracy of performance measure, to our knowledge for the first time.

II. EXPERIMENT 1

A. Overview

To investigate stream segregation in CI listeners using a temporal-discrimination task, experiment 1 employed the same task and stimulus sequence configurations used by [Roberts et al. \(2002\)](#) and also used ABA tone triplets. In contrast with [Hong and Turner \(2006\)](#), the stimuli were presented via direct input to the listeners' speech processors. The frequencies of the tones used were such that they fell in the center of the passband for the intended electrodes (see [Table I](#)). The speech processor was programmed so that it could deliver stimulation on only one channel at a time using the advanced combination encoders (ACE) strategy and selecting only one spectral peak. The precise pattern of electrical stimulation was verified by routing the stimuli through an experimental speech processor and by analyzing the output using a "dummy" implant within the manufacturer's com-

TABLE I. Frequency characteristics of the tones used in experiments 1 and 2 and their relation to the implant's 22 channels. Lower- and higher-numbered electrodes correspond to more basal and more apical places, respectively. Rows shown in bold indicate the restricted set of electrodes used in experiment 1. All electrode numbers were used in experiment 2.

Channel/electrode number	Lower frequency boundary (Hz)	Upper frequency boundary (Hz)	Channel center 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

puter interface. This allowed a frame-by-frame listing of the output of the implant's transmitter coil and the generation of an "electrogram," which illustrates visually the output on each electrode over the selected time window. This procedure replaced checking with a test implant and oscilloscope.

All stimuli were presented at C level (maximum comfortable level) and were loudness balanced using a standard clinical method of "sweeping" stimulation across the elec-

trode array; listeners had no control over the volume or sensitivity. Thus, the relation between electrode separation and temporal discrimination was measured directly across a wide range of separations, both for the longer tone sequences and the short triplets, in a group of CI listeners all using the same type of Nucleus implant. An interaction between 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 function relating threshold to electrode separation for the longer tone sequences than for the triplets.

B. Method

1. Stimuli and conditions

There were two conditions, one using longer sequences of alternating tones A and B and the other using short ABA triplets. The longer sequences had an identical configuration to those used by Roberts *et al.* (2002) and by Hong and Turner (2006); each sequence lasted 2.4 s and consisted of 12 AB cycles of alternating pure tones. These sequences were long enough to allow considerable build-up in the tendency for stream segregation in NH listeners (see, e.g., Roberts *et al.*, 2008). The duration of each tone was 60 ms, including 10-ms onset and offset ramps, and the standard inter-tone interval was 40 ms, corresponding to an onset-to-onset time of 100 ms between consecutive tones. The ABA triplets were constructed in the same way; each triplet lasted 300 ms. The frequency characteristics of the sequences were identical in the two conditions. Table I illustrates the relationship be-

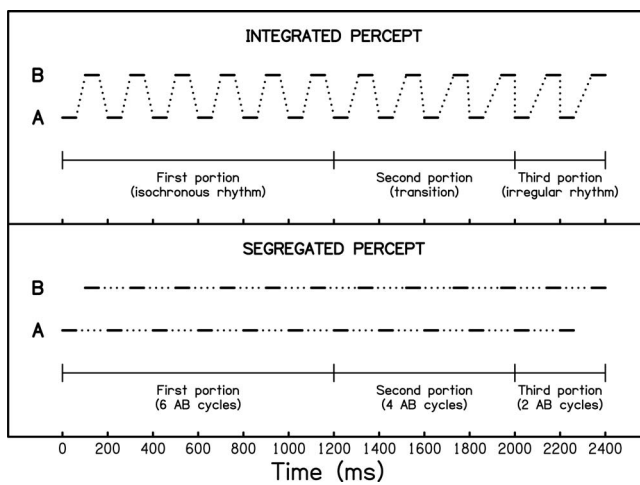


FIG. 1. Stimuli for experiment 1—schematic representation of the longer test sequences used. Each tone is depicted by a short solid line; tones heard as belonging to the same auditory stream are linked by dotted lines. The upper and lower panels represent a sequence heard as integrated (one stream) or segregated (two streams), respectively. Reproduced with permission from Roberts *et al.* (2002), p. 2078. Copyright 2002 by the Acoustical Society of America.

TABLE II. Demographic and other details for implant listeners in experiments 1 and 2.

Listener	Age	Gender	Expt.	Open-set speech recognition score (% correct for BKB sentences in quiet)	Type of electrode array	Pulse rate in normal daily use (pulses/s)	Pulse width in normal daily use (μ s)
L1	42	F	2	80	Perimodiolar	1200	25
L2	48	F	1	76	Straight	250	25
L3	61	F	2	91	Straight	1200	25
L5	58	M	1	48	Straight	250	25
L6	25	M	1, 2	98	Straight	250	25
L9	46	M	1	99	Straight	1800	25
L11	45	M	1, 2	92	Straight	900	25
L12	69	F	1	93	Perimodiolar	1200	25
L13	51	F	2	100	Straight	1200	25
L14	25	F	2	90	Straight	250	25

tween the frequencies of the tones used and the electrodes stimulated. Tone A was always at 1938 Hz and hence stimulated electrode 11 (e11). Tone B was at a frequency set to stimulate either one of six electrodes more basal than e11 (e1, e3, e5, e7, e9, or e10), or one of seven electrodes more apical than 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 every electrode was included for the region of most interest (centered on e11). The electrode used for tone B was selected quasi-randomly from the test set between trials but was constant within each sequence and did not vary within a trial.

In the longer-sequence condition, each tone B was presented at the exact midpoint in time between the preceding and following tones A for the standard (isochronous) sequences. The structure of the test (anisochronous) sequences is illustrated in Fig. 1. These sequences were configured in the same way as the standard case for the first six AB cycles, but tone B was delayed progressively in equal steps over the next four cycles, and the cumulated delay was maintained for the final two cycles. Hence, the rhythm was regular for the first half of a test sequence but changed to irregular in the second half. Cumulated delays of 10, 20, 30, or 40 ms were used in quasi-random order for the test sequences across trials; 40 ms was the maximum delay possible without temporal overlap between tones A and B. In the triplet condition, the test sequences were created by applying a delay of 10, 20, 30, or 40 ms to the middle tone B of the standard case. Thus, these test triplets were identical to the first three tones comprising the final two cycles in their longer counterparts. For all listeners, a constant pulse rate of 900 pps and a pulse width of 25 μ s were used; stimulation was monopolar (MP1+2), and pulses were interleaved across stimulated electrodes.

2. Listeners

Listeners were six adult, post-lingually deafened, experienced users of the Nucleus CI24 implant system; all had normal electrode insertions. Table II shows their demographic and other details.

3. Procedure

A two-interval, two-alternative, forced-choice procedure was used. In both conditions, all possible combinations of electrode pairings and onset delay for tone B were used in quasi-random order. Thus, there were four possible delays (10, 20, 30, or 40 ms) \times 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 combination of delay and electrode pair, giving a total of 112 presentations per block. A total of ten trial blocks was completed, giving 20 presentations of each combination.

All 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 supplied by the manufacturer for use with the Nucleus implant system. 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 reasonably 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."

The purpose of the experiment was explained to the listeners; they were instructed both verbally and in writing. Each listener was seated in front of a computer screen and keyboard. They were instructed to listen to both sequences of tones and to respond by pressing key 1 or 2 on the keyboard to indicate which sequence, first or second, was irregular in rhythm (i.e., to identify the anisochronous interval). Feedback was provided on the computer screen as to whether the response was correct or incorrect. All listeners received training and practice on both sequence lengths prior to the experimental runs. Stimulus pairs expected to be most easily discriminable were used, i.e., when tone B had the same frequency as tone A and the delay on tone B was either 0 or 40 ms. Pilot work indicated that all CI listeners performed well above chance in this case. Training was continued to ensure that each listener understood the task and what to listen for, responded confidently and appropriately, and

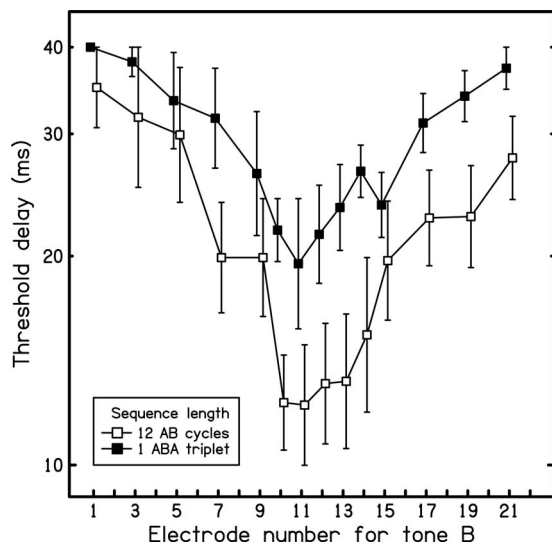


FIG. 2. Results for experiment 1—geometric mean threshold delays for six implant listeners in the triplets and the longer-sequence conditions (filled and open squares, respectively). For the calculation of means, thresholds above the maximum of 40 ms or below the minimum of 10 ms have been allocated values of 40 and 10 ms, respectively. Inter-subject standard errors for each mean are shown by vertical lines. Means and error bars for the two sequence lengths have been slightly displaced for clarity.

achieved good or near-perfect performance. Care was taken to ensure that any task learning was essentially complete before the experimental runs began.

C. Results

For each listener, the percentage of correct responses (out of 20) was calculated for each combination of delay and electrode separation. For each electrode separation, a threshold delay was derived from these scores using a logistic function to fit the data and to estimate the delay equivalent to an accuracy of 75% correct. This approach provided a good fit to the data in over 90% of cases. In cases of floor or ceiling effects on performance, thresholds of 40 or 10 ms were assumed, as appropriate.

The geometric mean thresholds for the six implant listeners are shown in Fig. 2. Overall, a strong effect of electrode number for tone B on threshold can be seen, with higher thresholds associated with greater electrode separation from e11, and this effect is broadly symmetrical. There is also clear evidence of an overall tendency toward higher thresholds for the triplets than for the longer sequences for all electrode positions. A within-subjects analysis of variance (ANOVA) performed using log-transformed threshold estimates 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 main effect of sequence length [$F(1,5)=7.652$, $p=0.040$], but there was not a significant interaction between electrode number and sequence length [$F(13,65)=1.104$, $p=0.372$]. To reduce the possibility that ceiling effects had distorted the outcomes of this analysis, the ANOVA was repeated, excluding the data for electrodes 1, 3, 19, and 21, which are toward each 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 [$F(9,45)=1.109$, $p=0.376$].

To explore further the relationship between threshold delay and electrode separation (between tones A and B) separately for the basal and apical directions (relative to e11), the ANOVA was repeated for each half of the electrode array. The data for e11 were included in both analyses. In the apical direction, there was a significant main effect of electrode separation [$F(6,30)=20.764$, $p<0.001$] and of sequence length [$F(1,5)=12.233$, $p=0.017$], but no significant interaction between sequence length and electrode separation [$F(6,30)=0.761$, $p=0.606$]. Indeed, the two curves are essentially parallel. In the basal direction, there was again a significant main effect of electrode separation [$F(6,30)=22.782$, $p<0.001$]. Although there was no significant main effect of sequence length [$F(1,5)=3.682$, $p=0.113$], a similar trend was apparent to that observed in the apical direction. Again, there was no significant interaction between sequence length and electrode separation [$F(6,30)=1.306$, $p=0.285$]. The hint of a narrowing distance between the two curves for the greatest electrode separations in the basal direction probably reflects ceiling effects on performance in the triplet condition. Excluding the more extreme electrode separations (e1 and e3 in the basal case; e19 and e21 in the apical case) did not change the outcome of these analyses.

Results for the six individual listeners are shown in Fig. 3. For the purposes of illustration, where threshold delays were greater than the maximum delay used (40 ms), a symbol has been plotted at the 40-ms point on each graph with an upward-pointing arrow. Similarly, where thresholds were less than the shortest delay used (10 ms), a symbol has been inserted at the 10-ms point with a downward-pointing arrow. Although 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 triplets and the longer sequences, with greater separations of tone B from e11 in both apical (higher electrode numbers) and basal directions (lower electrode numbers) generally associated with higher threshold delays. Also, for most listeners, thresholds for the triplets were generally worse than those for the longer sequences.

There were, however, some marked differences in performance between individuals. For example, L2 showed thresholds above 40 ms for the longer sequences on all the electrode separations across the array, except when tone B stimulated one of three central positions (e10, e11, and e12), and a similar pattern was apparent for the triplets. In contrast, L11 had thresholds of 10 ms or better for the longer sequences on all the electrodes stimulated by tone B from e3 through e14, with only slightly higher thresholds for the most basal and for the four most apical electrodes tested. For the triplets, however, L11 generally showed much higher thresholds with a clear dependence on electrode separation (higher thresholds for greater separations from e11).

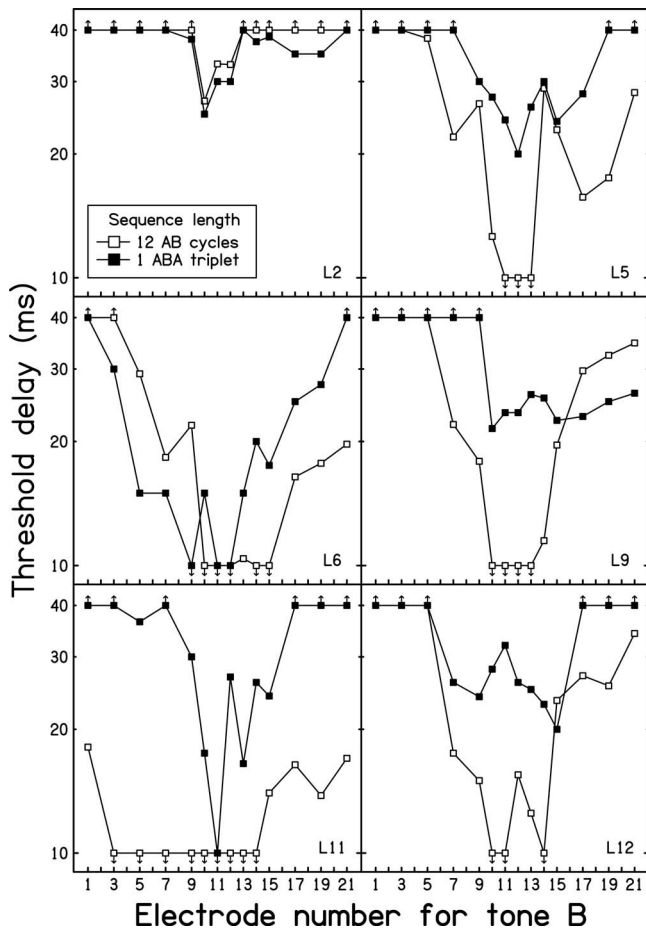


FIG. 3. Results for experiment 1—thresholds for six individual listeners in the triplets and the longer-sequence conditions (filled and open squares, respectively). An upward-pointing arrow on a symbol plotted at 40 ms indicates a threshold delay greater than the maximum tested. A downward-pointing arrow on a symbol plotted at 10 ms indicates a threshold delay lower than the minimum tested.

D. Discussion

The main findings can be summarized as follows. First, increased electrode separation between tones A and B is associated in all listeners with worse performance (higher threshold delays for tone B) for both longer sequences and triplets. Second, for most listeners, overall performance was generally worse for the triplets than for the longer sequences. Third, there was no interaction between sequence length and electrode separation.

The first of these findings is consistent with reports of temporal judgments for short sound sequences of different frequencies in NH listeners (e.g., van Noorden, 1975; Grose *et al.*, 2001) or which stimulate different electrodes in CI listeners (e.g., Hanekom and Shannon, 1998). The results demonstrate that CI listeners found judgments about the relative timing of sounds more difficult when they stimulated electrodes that were more widely separated spatially. The wider separation would have resulted in greater perceptual differences between the sounds, particularly in pitch or timbral brightness. The second finding, that performance was generally worse with the triplets than with the 2.4-s long sequences, tallies with anecdotal reports from our listeners that they found the triplet condition significantly more taxing

and is fully consistent with the findings of Hong and Turner (2006). While better performance overall in the triplet condition might have been expected as a result of a lack of build-up of stream segregation, any such benefit was clearly more than offset by the small number of tones available for listeners to detect the relative timing of tones A and B.

The third finding, that there was no interaction between electrode separation and sequence length, is very different from the results reported by Hong and Turner (2006). They measured thresholds when the ratios of the frequencies of tones A and B were 0 (1:1) and 0.5 (1:1.5). Thresholds for the latter case were reported as normalized values after dividing by the corresponding “baseline” performance in the former case; these normalized thresholds were typically much greater for the longer-sequence condition than for the triplet condition. This suggests a much stronger dependence of threshold on electrode separation for the longer sequences, presumably reflecting the contribution of stream segregation. The original thresholds from Hong and Turner’s (2006) data for the ratio of 0.5 can be reconstructed from the normalized values and the corresponding thresholds for the ratio 0. For the three individuals that they tested, these values indicate that the difference in thresholds between the triplets and longer sequences was usually much smaller for the ratio of 0.5 than for the ratio of 0, and it was sometimes reversed (i.e., crossovers were observed).

With respect to e11 (centered on 1938 Hz), frequencies giving a ratio of 0.5 in the present experiment would be 2907 Hz (basal direction) and 1292 Hz (apical direction). Table I indicates that these values would be produced by positioning tone B on e8 or e14, respectively. Given that the present experiment included comparisons of thresholds for the two sequence lengths across a much larger range of frequency separations than ratios of 0 and 0.5, one would expect to see crossovers in both the basal and apical directions for most, if not all, individual listeners. Typically, our listeners did not show any crossovers in their curves of threshold delay vs electrode separation (aside from L6 for electrodes 1–9 and L9 for electrodes 13–21). While it is not possible to prove a null hypothesis, it is worth noting that our ANOVA had sufficient power to show significant effects for electrode separation and sequence length. Indeed, as noted earlier, our mean results show broadly parallel curves for the triplets and longer-sequence conditions in both the basal and apical directions. Overall, these results clearly differ from those of Hong and Turner (2006).

Another issue that merits comment is our use of a log time scale to analyze the threshold estimates, following the approach used by Roberts *et al.* (2002, 2008) and by Hong and Turner (2006). If we had used a linear time scale instead, it is possible that a significant interaction between electrode separation and sequence length would have emerged. However, this would have been in the opposite direction to that predicted (i.e., steeper slopes for the triplets than for the longer-sequence condition). It should also be acknowledged that we cannot rule out entirely the possibility that CI listeners may require far longer sequences than do NH listeners (i.e., much greater than 2.4 s) for substantial build-up to occur. Note, however, that a much slower rate of build-up is

likely to reflect a greatly reduced tendency for stream segregation, most probably arising from the reduced effective separation between stimulation at different frequencies for CI listeners compared with NH listeners.

In summary, the results of experiment 1 did not provide convincing evidence that CI listeners commonly experience involuntary stream segregation. If the worse performance associated with greater electrode separations were providing an indirect measure of increased involuntary stream segregation, then a stronger dependence of threshold on electrode separation would have been expected for the 2.4-s long sequences than for the triplets. The latter were too brief to have allowed any appreciable build-up in the tendency for stream segregation to have occurred, and yet slopes for the function relating electrode separation to threshold delay were seen that were similar to those obtained with the longer sequences.

The results for experiment 1 support the conclusions of Cooper and Roberts (2007) rather than those of Hong and Turner (2006). To explore the ability of CI listeners to achieve voluntary stream segregation, for which selective attention is an important factor, experiment 2 used a task requiring the selection of a target pattern from a background of distractors.

III. EXPERIMENT 2

A. Introduction

An interleaved melody task (Dowling, 1973; Hartmann and Johnson, 1991; Cusack and Roberts, 2000) offers a potentially useful approach to investigating stream segregation in CI listeners (with the caveats outlined below), but thus far no published results are available from experiments of this type. The ability of CI listeners to recognize 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 six CI and six NH listeners. They pre-selected a set of 12 familiar songs and presented them on a musical synthesizer in two conditions, with or without rhythmic information (the latter case contained only pitch information—the notes were of equal duration with gaps between notes of equal duration). The NH listeners achieved near-perfect performance in melody recognition in both the rhythm and no-rhythm conditions. In contrast, the CI listeners' performance was around chance in the no-rhythm condition and above chance but significantly poorer than that for the NH listeners in the rhythm condition.

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 poor at melody recognition unless rhythmic cues are available. Thus, in the absence of rhythmic cues, target melodies used in such tasks with CI listeners should be very simple, comprising few elements. Pitch differences between successive notes should be as salient as possible, bearing in mind the generally poor pitch perception that CI listeners can achieve. Indeed, the percepts conveyed by place of electrical stimulation via an implant may only be

defined loosely as pitch; timbral brightness is arguably a better descriptor (Moore and Carlyon, 2005). If discrimination between two melodies is required, the difference between them should be clear and obvious to the participants.

CI listeners can hear pitch differences when different electrodes are stimulated, so they might be expected to be able to select and recognize a simple melody from a background of distracting tones, provided that the distractors sound sufficiently different from the tones contained in the melody. This outcome would provide evidence that CI listeners can use schema-based selection to hear out a subset of acoustic elements from a sequence as a separate stream, even in the absence of primitive stream segregation. Directing interleaved distractor tones to electrodes spatially separated along the array from the target tones should provide pitch (or brightness) differences that can be used to segregate the melody. If CI listeners can use such cues, their melody identification should be better when the distractors stimulate a different and distinct part of the electrode array than when they overlap with the part of the array stimulated by the target tones. Also, performance should be better when a loudness difference cue is additionally available. Note that any improvement seen when distractors are reduced in level cannot be accounted for by changes in energetic masking because distractor and target tones are presented at different times from one another.

In summary, two main predictions were tested in this experiment. First, in a task requiring discrimination between two simple tone sequences (melodies) interleaved with distractor tones, performance should be better when the distractors stimulate a spatially separate part of the electrode array from the melodies, compared with when the electrode ranges for the melodies and distractors overlap. Second, better performance should be achieved when the loudness of the distractors is reduced relative to that of the melody tones, compared with when they are at the same loudness.

B. Method

This experiment was a simplified version of the interleaved melody identification task used by Cusack and Roberts (2000). Two simple “melodies” were created, each a sequence of five pure tones, that were easy to distinguish from one another and to memorize for most CI listeners. Other details of the experimental set-up, including the configuration of the speech processor, frequency allocations, the mode of stimulation, and pulse characteristics, were identical to those used in experiment 1. The effect on recognition of interleaving the notes of the melody (target tones) with distractor tones was investigated for different target-distractor electrode separations and for different distractor levels, relative to the targets.

1. Stimuli and conditions

Each melody was centered on the middle of the electrode array; e9, e11, and e13 were stimulated. A spacing of two electrodes was used between successive sounds in an effort to provide a clear pitch (or brightness) contrast between them. The sequence of stimulation for melody 1 was

e9, e11, e13, e11, and e9. This should lead to a falling pitch followed by a rise, as higher-numbered electrodes are located toward the apical end of the electrode array and therefore should produce a lower pitch. For melody 2, the order was e13, e11, e9, e11, and e13, thus giving a rising pitch followed by a fall. No rhythm cues were available because all tones were equal in duration and were presented at the same rate.

The target tones comprising the melodies were interleaved with distractor tones, selected from one of five sets corresponding to different electrode ranges. Distractor sets whose ranges of electrode stimulation were spatially remote from the central portion of the array (which encompassed the target tones) were termed “distant,” and sets whose ranges were adjacent to but did not overlap with the targets were termed “neighboring.” The five sets and associated ranges were as follows: (i) distant-basal (DB), corresponding to e1–e5; (ii) neighboring-basal (NB), corresponding to e4–e8; (iii) overlapping (OV), corresponding to e9–e13 (same range as for the target tones); (iv) neighboring-apical (NA), corresponding to e14–e18; and (v) distant-apical (DA), corresponding to e18–e22. On each trial, six distractors from the appropriate range were selected randomly, in steps of one-electrode spacing, and were interleaved with the five targets such that the whole sequence (11 tones in total) began and ended with a distractor. This ensured that listeners could not perform the task correctly simply by listening for the first or last note.

As well as varying the region along the electrode array that was stimulated by the distractor tones, their level was systematically varied relative to that of the target tones. Without careful loudness-balancing of all the stimuli relative to one another, it is not possible to be sure of their precise relative loudness. Hence, the loudness of the distractors was expressed in terms of percentage of dynamic range. The threshold (T) and maximum comfortable loudness (C) levels for each electrode were measured using standard clinical methods before running the experiment. Stimulus levels are described in terms of stimulus units, which correspond to log current level. The target tones comprising the melody 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 and C levels were 100 and 200 stimulus units, respectively, the dynamic range would be 100 stimulus units. In this case, the levels of stimulation on each electrode would be 100% DR (100% of dynamic range) = 200 stimulus units, i.e., C level; 75% DR = 175 stimulus units; 50% DR = 150 stimulus units; and 25% DR = 125 stimulus units.

Individual stimuli comprising both the target melody and distractor sequence were pure tones that were 120-ms long, including 10-ms rise and fall times. The tone onset-to-onset time in an interleaved sequence was 200 ms; i.e., there was a silent interval of 80 ms between consecutive tones. The total duration of an interleaved sequence was 2.2 s. As before, the pure tones used were at frequencies equal to the center frequencies of the channel allocations defined in the experimental speech processor (see Table I). Figure 4 shows

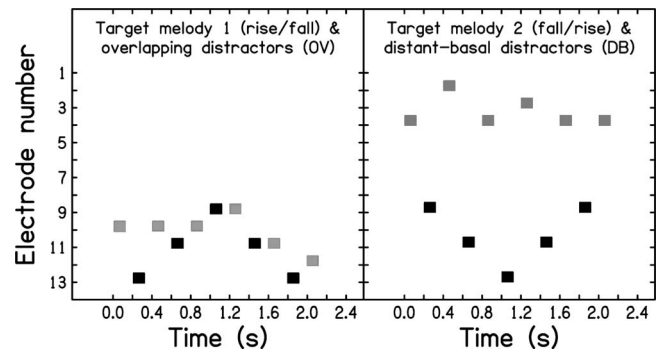


FIG. 4. Stimuli for experiment 2—electrodograms showing examples of interleaved sequences. The left panel shows target melody 1 (rise/fall in pitch) interleaved with distractor tones in the OV condition, i.e., taken at random from within the same electrode range as the target tones (e9–e13). The right panel shows target melody 2 (fall/rise in pitch) interleaved with distractor tones in the DB condition, i.e., taken at random from the range e1 to e5. The filled rectangles illustrate when electrical stimulation was present on each electrode. Black boxes show the target stimulation; gray boxes show the distractors. The ordinate shows electrode number, and the abscissa shows time elapsed in seconds. Only the range of electrode numbers that encompasses those stimulated is illustrated.

representative examples of electrodograms depicting sequences comprising the target tones of a melody interleaved with distractors.

2. Listeners

Listeners were six experienced adult CI users of the Nucleus CI24 device, all with normal electrode insertions and pitch ranking. Demographic and other details are shown in Table II. All listeners were screened to ensure that they were able to discriminate reliably between the two target melodies when presented in isolation (i.e., without distractors). A criterion was set of obtaining 100% correct on at least two successive experimental blocks (ten repetitions per block) following practice runs. Several potential participants screened in this way were rejected as they were unable to perform above chance for the isolated melodies, even after training.

3. Procedure

The experiment used a single-interval, two-alternative, forced-choice design. In the main experiment, the stimulus presented on each trial contained one or other of the target melodies (1 or 2), selected at random and interleaved with distractors. 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%, 50%, 75%, and finally 100% DR. In the 100%-DR case, the distractors were presented at approximately equal loudness to the target tones. Each block contained 100 presentations in a quasi-random order (5 distractor sets \times 2 target melodies \times 10 repetitions), giving a total of 400 trials in the main experiment (100 presentations \times 4 distractor levels). Every repetition of a block had an equal number of presentations of each melody, and a new randomization was used for the set of stimuli tested.

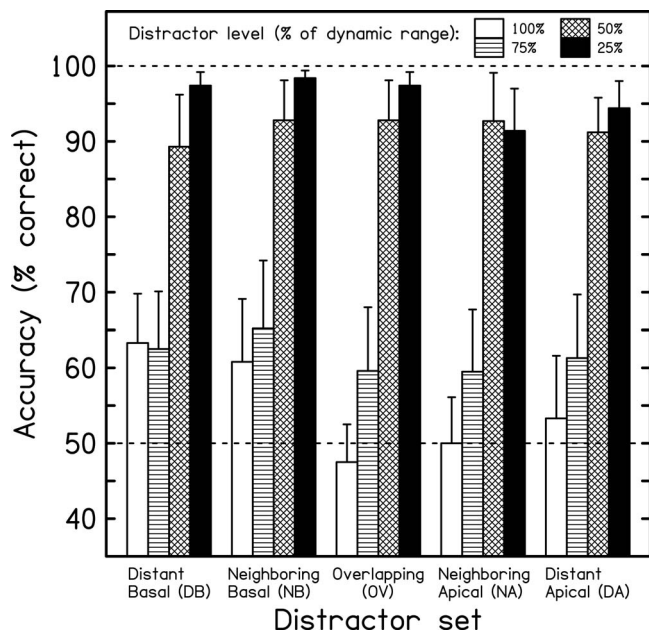


FIG. 5. Results for experiment 2—mean accuracy for six implant listeners. Clustered columns show results for each of the five distractor sets. The results for the four different distractor levels are shown by the differently shaded columns (see inset). Inter-subject standard errors for each mean are shown by vertical lines. The horizontal dashed lines indicate chance (50%) and perfect performance.

The purpose of the experiment was explained to the listeners; they were instructed both verbally and in writing. Listeners first received training in recognition of each melody in the absence of any distractors and then practice on the interleaved melody task prior to the main experimental runs. Care was taken to ensure that all CI listeners understood the task, that they were able to respond appropriately, and that any task learning was essentially complete before the experimental runs began. Each listener was seated in front of a computer screen and keyboard. A visual representation of each melody was provided on screen as a reminder of what to listen for. This illustrated the expected pitch changes for each melody. Listeners were instructed to respond after each trial by pressing either “1” or “2” on the keyboard to indicate which of the two target melodies they had heard. Feedback was provided on screen following each response, displaying “correct” or “incorrect” as appropriate.

C. Results

Responses in each condition were analyzed in terms of percentage correct responses; there was no evidence of bias toward reporting one melody in preference to the other. Mean results for all six CI listeners are shown in Fig. 5. 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% DR was 55%, 62%, 92%, and 96% correct, respectively. Note that overall performance was close to chance for the 100%- and 75%-DR conditions but very good for the 50%- and 25%-DR conditions.

Contrary to expectation, which distractor set was used appeared to have little effect on performance. In particular,

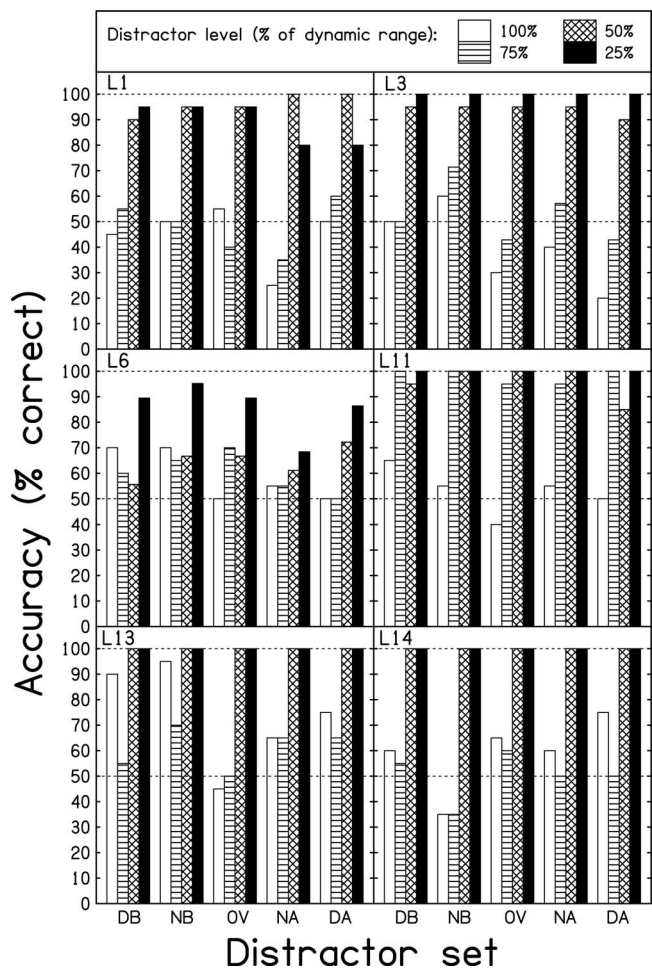


FIG. 6. Results for experiment 2—accuracy scores for six individual listeners. Clustered columns show results for each distractor set; abbreviations for these sets are as indicated in Fig. 5. The results for the four different distractor levels are shown by the differently shaded columns (see inset). The horizontal dashed lines indicate chance (50%) and perfect performance.

there was no consistent evidence of poorer performance in the OV condition. A within-subjects ANOVA showed a highly significant effect of distractor level [$F(3,15) = 16.56, p < 0.001$], but there was no effect of distractor set [$F(4,20) = 1.29, p = 0.307$], and there was no interaction between distractor set and level [$F(12,60) = 1.15, p = 0.338$]. As a precaution, to reduce heterogeneity of variance associated with percent correct scores, the ANOVA was repeated using arcsine-transformed data [$Y' = 2 \arcsin(\sqrt{Y})$; see [Keppel and Wickens \(2004\)](#), p. 155]. The results confirmed the outcome of the original analysis. Pairwise comparisons between the means for different distractor levels were calculated using the restricted least-significant-difference test ([Snedecor and Cochran, 1967](#)). Four out of the six possible comparisons showed a significant difference (100% vs 50% DR: $p = 0.004$; 100% vs 25% DR: $p = 0.001$; 75% vs 50% DR: $p = 0.023$; 75% vs 25% DR: $p = 0.005$). The difference between the 100%- and 75%-DR cases was not significant ($p = 0.472$), and the difference between the 50%- and 25%-DR cases was marginal ($p = 0.053$).

Individual results for the six CI listeners are shown in Fig. 6. Overall performance on this task in the 100%-DR condition was generally close to chance, with no systematic

difference between the different distractor sets. The only clear exception was L13; she showed the best scores for the two basally located distractor sets, DB and NB, both well above chance (90% and 95% correct, respectively), and the worst score for the OV set. Surprisingly, she performed better overall in the 100%-DR than in the 75%-DR condition. This may reflect some form of cue learning, as the 100%-DR condition was the last block run. Again, all but one of the CI listeners (in this case, L11) showed performance close to chance in the 75%-DR condition; whereas L11 showed strikingly good performance, with 100% correct scores for three of the five distractor sets and 95% correct for the other two (including the OV set). In general, the best performance was seen for the two lowest levels of distractor, i.e., the 50%- and 25%-DR conditions.

D. Discussion

With some exceptions, individual performance on this task was broadly consistent. Aside from one individual (L13) for some of the distractor sets, CI listeners were not able to identify reliably the target melodies when they were interleaved with distractors of equal loudness (i.e., 100% DR); they responded at or near chance, irrespective of the electrode range occupied by the distractors. When the distractors were attenuated to 75% DR, performance was at (or only slightly above) chance for five out of six listeners. However, further attenuation of the distractors improved performance considerably. At 50% DR, five out of six listeners were able to demonstrate near-perfect performance, and when the distractors were reduced to 25% DR, all listeners showed above-chance (and usually near-perfect) performance. Although they do not provide any details, Chatterjee *et al.* (2006) claimed on the basis of their pilot work that loudness differences can be used to segregate sequences of stimuli.

The fact that the mean results for each distractor set did not differ significantly from each other shows that these CI listeners were not generally able (with the possible exception of L13) to benefit from any pitch (or brightness) differences between the target tones and distractors. In principle, the inclusion of intermediate distractor levels (between 75% and 50%) might have allowed us to distinguish more effectively between the impacts of different distractor sets. However, even if this were true, a much smaller effect of distractor set is implied than was anticipated. In contrast, three NH listeners who completed a comparable task showed a strong dependence of performance on distractor set. These listeners were presented with pure-tone targets and distractors via headphones (Sennheiser HD535) at equal levels (65 dB SPL) in a sound-attenuating room. The frequencies of the tones corresponded to those used for the CI listeners (see Table I), although it should be noted that the pitches evoked will have differed between the NH and CI listeners owing to place mis-matching that arises from the restricted insertion depth characteristic of cochlear implantation. So long as the distractors were drawn from a different frequency range to the target tones, all three NH listeners scored well above chance (two with near-perfect performance), even when the distractors were equal in level to the targets.

Thus, the performance of our CI listeners on this task differs substantially from that obtained with NH listeners; of the two predictions given earlier, only the second has been upheld. Based on these results, it would seem that CI listeners are generally not able to utilize 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 targets. This outcome appears to be inconsistent with Chatterjee and Galvin's (2002) claim that electrode separation influences the ability of listeners to hear a preview rhythm embedded in a larger sequence, as that implies a form of pattern recognition similar to that needed to identify successfully an interleaved melody. The reason for this discrepancy is unclear in the absence of detailed information about their study, but it may relate to their use of subjective ratings, in contrast with our use of an accuracy measure. Nonetheless, our CI listeners were able to ignore the distractors and successfully attend to the target melodies, provided that the distractors were reduced in level on each electrode to 50% DR or below. This would be equivalent to attending to sounds at the maximum comfortable level while ignoring distracting sounds that are clearly audible but softer. Of course, this finding does not necessarily imply that CI listeners would be able to use the level difference to attend to targets presented at 50% DR when accompanied by distractors presented at C level, as the relative audibility of the targets is also likely to be an important factor affecting performance.

IV. GENERAL DISCUSSION

The results of experiment 1 suggest that our CI listeners generally did not experience involuntary stream segregation when making judgments of the relative timing of tones of different frequency. This task was intended to provide an objective 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 separation between successive tones led to worse performance on judgments about their relative timing, presumably as a result of perceptual differences between them (i.e., differences in pitch or brightness). However, this effect was found for short sequences (tone triplets) to about the same degree as for the longer sequences, and it is therefore not possible to conclude that the effect of electrode separation on temporal discrimination was mediated by stream segregation.

This outcome contrasts with the findings of Hong and Turner (2006). However, it is consistent with those of Cooper and Roberts (2007), who found that judgments by CI listeners of whether a tonal sequence is heard as one stream or two do not conform to the pattern seen in NH listeners. Specifically, the effect of electrode separation on streaming judgments did not depend on sequence rate and did not show evidence of spontaneous flipping between alternative interpretations (Anstis and Saida, 1985). Chatterjee *et al.* (2006) did not test the effect of sequence rate in their study of auditory streaming in CI listeners. Cooper and Roberts (2007) suggested that the perceptual space within which successive sounds can differ from one another is much more limited for

CI listeners than for NH listeners and that this in turn impairs the ability of CI listeners to achieve stream segregation (Moore and Gockel, 2002). Cooper and Roberts (2007) concluded that the judgments of their CI listeners were more akin to a measure of channel discrimination (i.e., a measure of how many pitches were heard) than to a genuine measure of stream segregation.

When there are no effective cues for involuntary stream segregation, CI listeners might, in principle, rely instead on schema-based selection to hear out a subset of acoustic elements from a sequence as a separate stream. Such a reliance on effortful attentional mechanisms would inevitably limit the ability of CI listeners to cope with complex listening environments, particularly given the fairly impoverished nature of the sensory information they receive from their implants. It is, therefore, noteworthy that stimulation of electrodes in differing portions of the electrode array generally did not enable CI listeners in experiment 2 to ignore the distractors and attend to the target melody. At the very least, one might have expected the benefits of differences in electrode range and in level to have combined to enable schema-based selection of the target melody from the distractors, but no evidence of such an interaction was found.

The apparent absence of an effect of distractor set on interleaved melody recognition suggests that that the perceptual differences between sensations evoked by stimulating different places along the electrode array may not be great enough even to support schema-based selection. This may seem somewhat surprising, given that our CI listeners were generally aware of pitch differences between stimulation on different electrodes and were able to discriminate with almost perfect accuracy between our two simple melodies, in the absence of any distractors. However, in NH listeners, the frequency separation at the FB is usually about eight times larger than the frequency difference limen (FDL) below 2 kHz (Rose and Moore, 2005), indicating that the ability to discriminate between two sounds is not in itself sufficient to enable stream segregation by selecting voluntarily a subset of tones from a larger sequence. Nonetheless, it must be acknowledged that the relationship between stream segregation and frequency discrimination is a complex one. In particular, a much wider range of FB/FDL ratios is typical of hearing-impaired listeners, and even in NH listeners this ratio declines as the frequency is raised beyond 2 kHz (Rose and Moore, 2005).

In summary, 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 most CI listeners, either for involuntary or schema-based segregation. Stimulation of different electrodes does produce perceptual differences, but it appears that these are generally insufficient to facilitate the perceptual grouping of sounds. Differences in level can be used to select target tones interleaved with distractors, but even these must be large to obtain above-chance performance in most CI listeners. On the basis of these results, and of related studies of simultaneous grouping in CI listeners (Carlyon *et al.*, 2007; Cooper, 2008), one might speculate that the ability of many CI listeners to perform well in speech recognition

tasks has very little to do with auditory grouping. The processing strategies used in implants are optimized for speech recognition. For example, the Nucleus devices used by all our listeners employ the ACE strategy (n of m channels), which leads to the activation of only a subset of electrodes in the array, specifically those associated with frequency bands containing the largest spectral peaks. This approach is a robust way of representing the critical formant frequencies for successful speech perception except under very adverse signal-to-noise ratios.

ACKNOWLEDGMENTS

This research was supported by the “Hear and Now” Trust Fund (University Hospital Birmingham, U.K.). We are grateful to Brian Moore and to Rob Morse for their comments on an earlier version of this manuscript. Preliminary presentations of this research were given at the following meetings: (a) British Society of Audiology, Short Papers on Experimental Studies of Hearing and Deafness (University College London, September 2004); (b) Conference on Implantable Auditory Prostheses (Asilomar, Pacific Grove, CA, July–August 2005); (c) British Cochlear Implant Group (BCIG) Meeting (Dublin, Ireland, April 2007).

¹The dependence of gap detection thresholds on channel separation is reduced for most CI listeners by extensive training (van Wieringen and Wouters, 1999), but it is not abolished.

- Anstis, S., and Saida, S. (1985). “Adaptation to auditory streaming of frequency-modulated tones.” *J. Exp. Psychol. Hum. Percept. Perform.* **11**, 257–271.
- Bey, C., and McAdams, S. (2002). “Schema-based processing in auditory scene analysis.” *Percept. Psychophys.* **64**, 844–854.
- Bey, C., and McAdams, S. (2003). “Postrecognition of interleaved melodies as an indirect measure of auditory stream formation.” *J. Exp. Psychol. Hum. Percept. Perform.* **29**, 267–279.
- Bregman, A. S. (1978). “Auditory streaming is cumulative.” *J. Exp. Psychol. Hum. Percept. Perform.* **4**, 380–387.
- Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT, Cambridge, MA).
- Bregman, A. S., and Campbell, J. (1971). “Primary auditory stream segregation and perception of order in rapid sequences of tones.” *J. Exp. Psychol.* **89**, 244–249.
- Carlyon, R. P., Long, C. J., Deeks, J., and McKay, C. M. (2007). “Concurrent sound segregation in electric and acoustic hearing.” *J. Assoc. Res. Otolaryngol.* **8**, 119–133.
- Chatterjee, M., and Galvin, J. J. (2002). “Auditory streaming in cochlear implant listeners.” *J. Acoust. Soc. Am.* **111**, 2429.
- Chatterjee, M., Sarampalis, A., and Oba, S. I. (2006). “Auditory stream segregation with cochlear implants: A preliminary report.” *Hear. Res.* **222**, 100–107.
- Cooper, H. R. (2008). “Auditory grouping in cochlear implant listeners.” Ph.D. thesis, Aston University, Birmingham, UK.
- Cooper, H. R., and Roberts, B. (2007). “Auditory stream segregation of tone sequences in cochlear implant listeners.” *Hear. Res.* **225**, 11–24.
- Cusack, R., and Roberts, B. (2000). “Effects of differences in timbre on sequential grouping.” *Percept. Psychophys.* **62**, 1112–1120.
- Dowling, W. J. (1973). “The perception of interleaved melodies.” *Cognit. Psychol.* **5**, 322–337.
- Fu, Q.-J., and Nogaki, G. (2005). “Noise susceptibility of cochlear implant users: The role of spectral resolution and smearing.” *J. Assoc. Res. Otolaryngol.* **6**, 19–27.
- Grose, J. H., Hall, J. W., Buss, E., and Hatch, D. (2001). “Gap detection for similar and dissimilar gap markers.” *J. Acoust. Soc. Am.* **109**, 1587–1595.
- Hanekom, J. J., and Shannon, R. V. (1998). “Gap detection as a measure of electrode interaction in cochlear implants.” *J. Acoust. Soc. Am.* **104**, 2372–2384.

- Hartmann, W. M., and Johnson, D. (1991). "Stream segregation and peripheral channeling," *Music Percept.* **9**, 155–183.
- Hong, R. S., and Turner, C. W. (2006). "Pure-tone auditory stream segregation and speech perception in noise in cochlear implant recipients," *J. Acoust. Soc. Am.* **120**, 360–374.
- Keppel, G., and Wickens, T. D. (2004). *Design and Analysis: A Researcher's Handbook*, 4th ed. (Prentice-Hall, Englewood Cliffs, NJ).
- Kong, Y.-Y., Cruz, R., Jones, J. A., and Zeng, F.-G. (2004). "Music perception with temporal cues in acoustic and electric hearing," *Ear Hear.* **25**, 173–185.
- Moore, B. C. J., and Carlyon, R. P. (2005). "Perception of pitch by people with cochlear hearing loss and by cochlear implant users," in *Pitch, Neural Coding and Perception*, Handbook of Auditory Research Vol. **24**, edited by C. J. Plack, A. J. Oxenham, R. R. Fay, and A. N. Popper (Springer, Berlin).
- Moore, B. C. J., and Gockel, H. (2002). "Factors influencing sequential stream segregation," *Acta Acust. Acust.* **88**, 320–333.
- Roberts, B., Glasberg, B. R., and Moore, B. C. J. (2002). "Primitive stream segregation of tone sequences without differences in fundamental frequency or passband," *J. Acoust. Soc. Am.* **112**, 2074–2085.
- 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," *J. Exp. Psychol. Hum. Percept. Perform.* **34**, 992–1006.
- Rogers, W. L., and Bregman, A. S. (1993). "An experimental evaluation of three theories of auditory stream segregation," *Percept. Psychophys.* **53**, 179–189.
- Rogers, W. L., and Bregman, A. S. (1998). "Cumulation of the tendency to segregate auditory streams: Resetting by changes in location and loudness," *Percept. Psychophys.* **60**, 1216–1227.
- Rose, M. M., and Moore, B. C. J. (2005). "The relationship between stream segregation and frequency discrimination in normally hearing and hearing-impaired subjects," *Hear. Res.* **204**, 16–28.
- Snedecor, G. W., and Cochran, W. G. (1967), *Statistical Methods*, 6th ed. (Iowa University Press, Ames, IA).
- Throckmorton, C. S., and Collins, L. M. (2002). "The effect of channel interactions on speech recognition in cochlear implant subjects: Predictions from an acoustic model," *J. Acoust. Soc. Am.* **112**, 285–296.
- van Noorden, L. P. A. S. (1975). "Temporal coherence in the perception of tone sequences," Ph.D. thesis, Eindhoven University of Technology, The Netherlands.
- van Wieringen, A., and Wouters, J. (1999). "Gap detection in single- and multiple-channel stimuli by LAURA cochlear implantees," *J. Acoust. Soc. Am.* **106**, 1925–1939.
- 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," *J. Acoust. Soc. Am.* **106**, 938–945.
- Warren, R. M., Obusek, C. J., Farmer, R. M., and Warren, R. P. (1969). "Auditory sequence: Confusion of patterns other than speech and music," *Science* **164**, 586–587.