

Concurrent Sound Segregation in Electric and Acoustic Hearing

ROBERT P. CARLYON,¹ CHRISTOPHER J. LONG,¹ JOHN M. DEEKS,¹ AND COLETTE M. MCKAY²

¹*MRC Cognition & Brain Sciences Unit, 15 Chaucer Rd, Cambridge CB2 7EF, England*

²*School of Life and Health Sciences, Aston University, Aston Triangle, Birmingham B4 7ET, England*

Received: 11 July 2006; Accepted: 4 December 2006; Online publication: 10 January 2007

ABSTRACT

We investigated potential cues to sound segregation by cochlear implant (CI) and normal-hearing (NH) listeners. In each presentation interval of experiment 1a, CI listeners heard a mixture of four pulse trains applied concurrently to separate electrodes, preceded by a “probe” applied to a single electrode. In one of these two intervals, which the subject had to identify, the probe electrode was the same as a “target” electrode in the mixture. The pulse train on the target electrode had a higher level than the others in the mixture. Additionally, it could be presented either with a 200-ms onset delay, at a lower rate, or with an asynchrony produced by delaying each pulse by about 5 ms *re* those on the nontarget electrodes. Neither the rate difference nor the asynchrony aided performance over and above the level difference alone, but the onset delay produced a modest improvement. Experiment 1b showed that two subjects could perform the task using the onset delay alone, with no level difference. Experiment 2 used a method similar to that of experiment 1, but investigated the onset cue using NH listeners. In one condition, the mixture consisted of harmonics 5 to 40 of a 100-Hz fundamental, with the onset of either harmonics 13 to 17 or 26 to 30 delayed *re* the rest. Performance was modest in this condition, but could be improved markedly by using stimuli containing a spectral gap between the target and nontarget harmonics. The results suggest that (a) CI users are unlikely to use temporal pitch differences between adjacent chan-

nels to separate concurrent sounds, and that (b) they can use onset differences between channels, but the usefulness of this cue will be compromised by the spread of excitation along the nerve-fiber array. This deleterious effect of spread-of-excitation can also impair the use of onset cues by NH listeners.

Keywords: cochlear implants, grouping, pitch, onset delays

INTRODUCTION

Although many cochlear implant (CI) users understand speech well in quiet, performance is severely degraded in the presence of competing sounds (Skinner et al. 1994; Stickney et al. 2004). Some insight into the reasons for this poor performance has come from speech studies that manipulated the nature of the interfering sounds. For example, when the masker is a noise, then, unlike normal-hearing (NH) listeners, CI users do not perform better when the masker is amplitude-modulated at rates between 2 and 8 Hz, compared to when it is steady (Nelson et al. 2003). This suggests that, over a range of modulation rates, they are unable to exploit portions of the stimulus where the signal-to-masker ratio is momentarily high. In addition, when the masker is a single talker, CI patients do not benefit from differences between the sex of the target and masker (Stickney et al. 2004), suggesting that they do not effectively exploit differences in fundamental frequency (“F0”) and/or vocal-tract length.

In contrast to speech studies such as those described above, there has been relatively little *psycho-physical* research on the cues that CI users can and

Correspondence to: Robert P. Carlyon • MRC Cognition & Brain Sciences Unit • 15 Chaucer Rd, Cambridge CB2 7EF, England. Telephone: +44-1223-355294; fax: +44-1223-359062; email: bob.carlyon@mrc-cbu.cam.ac.uk

cannot use for concurrent sound segregation. This is perhaps surprising, given that experiments using psychophysical techniques and/or highly simplified stimuli have been useful in refining our understanding of sound segregation by NH listeners. Such studies allow one to differentiate between processes related to the organization of sounds over time (e.g., auditory streaming) and those involved in assigning concurrent frequency components to the appropriate source (“concurrent sound segregation”) (Bregman 1990; Carlyon and Gockel *in press*). They also reveal not only which cues are most important for sound segregation, but also the conditions under which those cues will and will not be effective. For example, a review of the literature reveals that, for sounds that overlap in time, the most powerful cues for segregation are differences in onset time and in F0 between the two sources (Darwin and Carlyon 1995). However, onset differences do not appear to be useful when the two sources excite the same region of the cochlea (Carlyon 1996a, b), and F0 differences ($\Delta F0$ s) are relatively ineffective when conveyed by harmonics that are unresolved by the peripheral auditory system (Darwin 1992; Carlyon 1996a; Darwin and Carlyon 1995).

This article investigates the potential usefulness of two cues to concurrent sound segregation by CI users. One of these relates to differences in F0 between sources. NH listeners primarily extract F0 from the auditory nerve (AN) responses to individual, low-numbered harmonics, each of which excites a more-or-less discrete set of fibers. For such stimuli, the frequency of each harmonic could be conveyed by a temporal (phase locking) code, and/or by place-of-excitation. Furthermore, the phase-locking to a given frequency is conveyed by AN fibers tuned close to that frequency, a fact that may be important for the analysis of pitch by more central processes (Moore 1982). In contrast, most CIs encode F0 by the rate of envelope modulation applied to a fixed-rate pulse train on each channel. Such modulations do not convey place-of-excitation cues, and modulations of a few hundred Hertz may be conveyed by AN fibers tuned to frequencies of several thousand Hertz (Ketten et al. 1998; Oxenham et al. 2004). To investigate the use of this “purely temporal” code to F0, we studied simplified stimuli in which each stimulated channel of a CI conveyed a periodic pulse train consisting of a single pulse per period. Our results show that CI listeners cannot exploit pulse-rate differences between electrodes to perceptually segregate a pulse train applied to one channel of a four-channel stimulus.

The second cue that we studied relates to onset differences between sources. Unlike F0, there is no qualitative difference between the way onset time is

encoded by CIs compared to NH. Hence, a priori, one might expect onset cues to be more easily exploited by CI users. To test this idea, we required subjects to perceptually segregate a pulse train applied to one channel of a four-channel stimulus, in which the target channel started 200 ms later than, and stopped at the same time as, the others. Results obtained with five CI users showed that they could achieve a modest benefit from the onset cue. Experiment 2 studied the onset cue using analogous stimuli with NH listeners. It is concluded that the usefulness of onset cues depends strongly on the extent to which the target and nontarget channels excite discrete regions of the cochlea.

EXPERIMENT 1

Rationale

The choice of paradigm to be used to study concurrent sound segregation by CI listeners is somewhat constrained compared to those previously employed with NH listeners. One approach to studying segregation in NH has been to manipulate the influence of a component or subset of components on the pitch, timbre, or phonetic identity of a complex sound, by varying some physical parameter such as their onset times or frequencies (Bregman and Pinker 1978; Darwin 1981; Darwin and Sutherland 1984; Darwin and Gardner 1986; Darwin and Ciocca 1992; Hukin and Darwin 1995). Such an approach requires a good understanding of what the perceptual effect of removing those component or components would be. Unfortunately, we do not have such a priori knowledge when considering the effects of removing or segregating one channel of electrical stimulation.

A second approach, which has been used profitably in NH, is to require subjects to determine whether an isolated “probe” tone is also present as a component of a complex tone (e.g., Hartmann et al. 1990). We adapted this approach by requiring CI listeners to compare stimulation on a single channel of their implant to that applied concurrently to four channels. We then introduced manipulations, such as differences in onset times and pulse rate between the “target” and “nontarget” channels in the mixture, and investigated whether, by improving segregation of the target, these manipulations increased performance on the task. This increase was measured relative to a baseline task in which these additional cues were absent. Because we did not want our CI subjects to be faced with an impossible task in any condition, we increased the current applied to the target channel, relative to that of the others, in all conditions including the baseline.

Subjects

Five adult postlingually deafened users of the Nucleus CI24 implant took part. Their details are given in Table 1.

Experiment 1a: overview of conditions

Experiment 1a consisted of a “baseline” condition and several experimental conditions. The baseline condition is illustrated schematically in Figure 1a. Each half of each two-interval trial consisted of a 400-ms “mixture” preceded, 200 ms earlier, by a 400-ms “probe.” The two halves of the trial were separated by 2 s. The mixture consisted of four 100-pps-pulse trains, each presented concurrently to one of four electrodes. We will refer to the four electrodes as A, B, C, and D, where A is the most apical and D the most basal, as the electrodes used differed from subject to subject. Each electrode in the mixture except one was presented at the same percentage of the dynamic range (DR) for that electrode when presented alone. One electrode—the “target”—was presented at a higher percentage of its DR than the rest. This is indicated in Figure 1a by the “taller” pulses on channel B. The target electrode was always B or C and was the same for the two presentations in each trial. In one presentation in a trial the probe was identical to the pulse train presented on the target electrode in the mixture. This presentation was termed the “signal” interval. In the “standard” interval, the probe was the same as the pulse train applied to the other possible target electrode (i.e., C or B) and had the same level as on those trials in which that other electrode was the target.

It is worth pointing out two important features of the above design. First, although the target electrode was presented at a higher current level than the others, subjects could not perform the task simply by comparing the two mixtures on each trial, in a way analogous to “profile analysis” in NH (Green 1988),

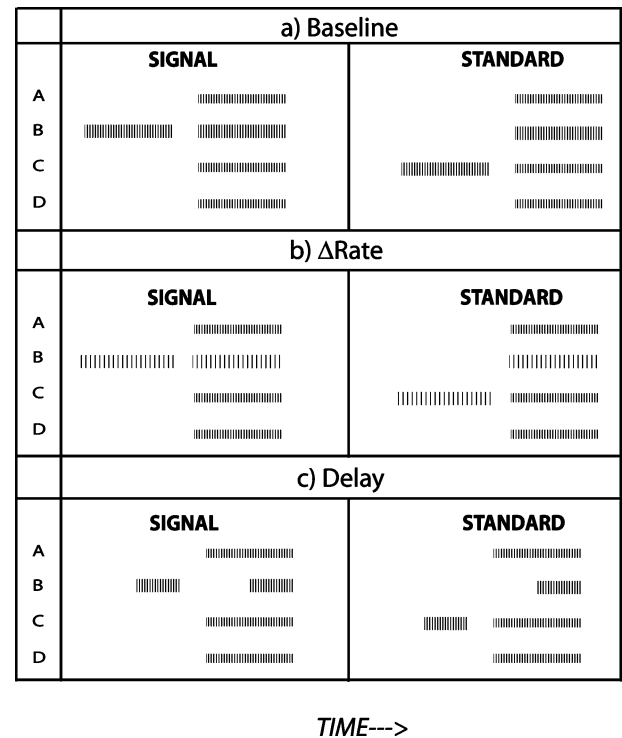


FIG. 1. (a) A schematic representation of the stimuli in two halves of a two-interval forced choice trial in the baseline condition of experiment 1, for a trial where the target is on channel B. The “taller” pulses of the target channel in the mixture represent their enhanced amplitude *re* those on the nontarget channels. Fewer pulses per stimulus are shown than were actually presented. (b) A representation of the stimuli in the Δ Rate condition. The difference in pulse rate between the target and nontarget channels has been exaggerated for clarity. (c) A representation of the stimuli in the Delay condition.

because the two mixtures were identical on each trial. Second, although the probes differed between the two halves of the trial, subjects could not perform the task by identifying the “signal” probe, because this differed from trial to trial. Rather, the subject had to compare the probe to one channel in the mixture: they were instructed to identify that interval where the probe was most clearly present in the subsequent

TABLE 1

Details of the CI users who took part in the experiments							
Subject ID	Age (years)	Etiology	HLD (years)	CI (years)	BKBq (%)	BKBn (%)	Channels (A,B,C,D)
CI1	75	Otosclerosis/noise	51	2	90	69	7,11,15,19
CI2	61	Unknown	22	1	98	87	8,12,16,20
CI3	46	Measles/unknown	29	2	75	41	8,12,16,20
CI4	56	Unknown	41	3	90	85	8,12,16,20
CI5	67	CSOM	>50	4	79	45	7,11,15,19

CI indicates the number of years since CI activation. BKBq indicates performance with Bench–Kowal–Bamford sentences (Bench and Bamford 1979) in quiet; BKBn indicates that with the sentences presented in pink noise at a signal-to-noise ratio of 10 dB. Each channel is specified in terms of the more apical electrode comprising the bipolar pair. The electrodes are numbered 1 to 22 in basal-to-apical direction.

HLD = hearing loss duration, BKB = Bench–Kowal–Bamford, CSOM = chronic serous otitis media.

mixture. A practice run was performed before each test condition, but no feedback was given during the tests.

The “Arate” condition is illustrated in Figure 1b. It differed from the baseline condition in that the pulse rate applied to the target electrode was reduced from 100 to 77 pps. We wished to determine whether this rate difference helped the listener to “hear out” the target in the mixture, thereby improving performance relative to the baseline condition. Because we wished to maximize the chances of subjects being able to exploit rate differences, we chose to introduce a difference between two quite low rates, where sensitivity in a *sequential* rate discrimination task is quite good, and to avoid higher rates where performance in sequential tasks often deteriorates (Shannon 1983; Tong and Clark 1985; Townshend et al. 1987; McKay et al. 2000; Zeng 2002). It is also worth noting that, at least for sequential tasks, discrimination of these low-rate pulse trains is better than that of the modulation rate applied to high-rate carriers, such as those used in some CI speech-processing strategies (Baumann and Nobbe 2004). By choosing stimuli that are discriminable in a sequential task, we hoped to maximize our chances of observing any effect of Δ rate in the segregation of concurrent stimuli. In this Δ rate condition, the current level on all channels was kept the same as in the baseline condition, because the function relating loudness to pulse rate is flat between 77 and 100 pps (McKay and McDermott 1998).

Figure 1c illustrates one trial in the “Delay” condition. It was identical to the baseline condition except that the pulse train presented to the target electrode started 200 ms after that applied to the others, and ended at the same time as them. We wished to determine whether this onset delay helped listeners to “hear out” the target in the mixture, thereby improving performance relative to the baseline condition. Because this reduced the duration of the pulse train on the target channel from 400 to 200 ms, the duration of the probe was also reduced to 200 ms.

Two further experimental conditions were included. As Figure 2 shows, introducing a rate difference also introduces a large and time-varying asynchrony between the pulses in the target and nontarget channels. To determine whether any effect of the rate difference was due to this asynchrony, we included the “Asynch” condition, in which the pulses in the target channel were delayed by 5 ms (Fig. 2). We also included an “AsynchDelay” condition, which was a combination of the “asynchrony” and “delay” cues. It was identical to the Asynch condition, except for the introduction of a 200-ms onset delay and reduction of the probe duration to 200 ms.

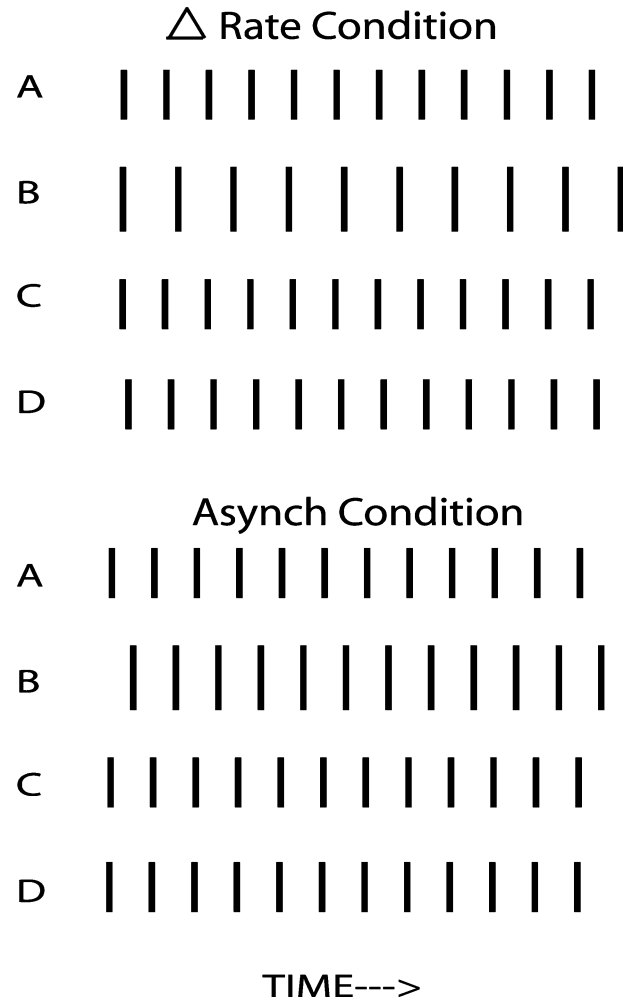


FIG. 2. Schematic of the timing of the pulses in the different channels of the mixture in the Δ rate and Asynch conditions of experiment 1a. Fewer pulses per stimulus are shown than were actually presented. The example shows a case where the target was on channel B.

Each biphasic pulse in each channel consisted of two 100- μ s phases of opposite polarity separated by a gap of 43 μ s. All presentation was in so-called “BP+1” mode, with the return electrode being separated from the stimulating electrode by a single “unused” contact. The timing of pulses within each period in the baseline condition is shown schematically in Figure 3, for the case where the target was on channel B. There was a 50- μ s gap between subsequent pulses, and, in all conditions, the order was basal to apical except that the last pulse in each period was always on the target channel. The four pulses were repeated every 10 ms, to give a rate of 100 pps per electrode, except for the target channel in the Δ rate condition, where the rate was 77 pps. To check that subjects were not performing the task by attending to the very last pulse in the stimulus, which was also always on the target channel, we

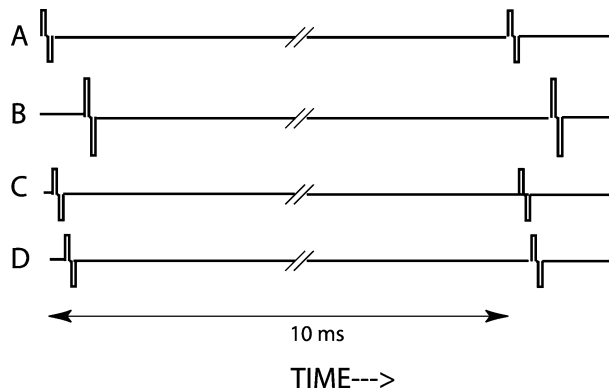


FIG. 3. Schematic showing a close-up of the relative timing of the pulses in the different channels comprising the mixture in the baseline condition for a trial in which the target was on electrode B.

included an additional baseline condition, “BaseDrop,” in which this final pulse was dropped.

Experiment 1a: preliminary measures

Prior to the start of the experiment, we selected the subset of electrodes to be tested for each subject. This choice was guided by the requirements that none of the electrodes in the set had abnormal impedances or were deactivated in clinical use for any other reason. The positions of electrodes A, B, C, and D are indicated separately for each subject in the last column of Table 1. The spacing between electrodes in the Nucleus CI24 device is 0.75 mm. Because each electrode was separated from its nearest neighbor by four electrode positions, the spacing between adjacent electrodes was 3 mm.

Once the electrode set had been selected, pretests confirmed that pulse trains on electrodes B and C were easily discriminable when presented on their own. Threshold (“T”) and most-comfortable (“C”) levels were obtained for each electrode and are shown in Table 2a. Thresholds were obtained using a two-interval forced-choice task and a 2-up 1-down adaptive procedure (Levitt 1971). Thresholds were estimated from the mean of the last eight turn-points in the procedure, for which the step size was five clinical units (CUs). C levels were obtained by presenting the stimulus at a moderate level and then increasing the level on subsequent presentations until it reached the maximum level that the subject said he/she would be comfortable listening to intermittently for 2 or 3 h.

We then performed a set of pilot measures to identify the levels to be used for each electrode and subject. The current level on the targets and nontargets in a mixture were set so that they did not exceed a comfortable level when presented together. We also aimed to choose levels that produced

performance which, when averaged across the two target electrodes, was neither at ceiling nor at chance. The number of measures needed to converge on this solution differed across subjects, but in all cases, the following procedures were adopted: (1) The current levels applied to each electrode of a four-electrode mixture were set to the same percentage of the subject’s DR for that electrode alone, defined as C–T in linear microamperes. These levels were then covaried using a loudness-balancing procedure (see below) so that the mixture’s loudness was the same as that of electrode B presented at 65% of DR. The resulting value, termed $x\%$ DR, was always less than 65 due to loudness summation across electrodes. (2) We then confirmed that electrodes B and C, presented at 65% of DR, had equal loudness. (3) The stimulus level for each target electrode was initially set to 70% DR, and that of the nontarget electrodes to $x\%$ DR. However, these target and nontarget levels sometimes had to be adjusted in order to avoid excessive loudness and/or to ensure that performance in the baseline condition was between chance and ceiling. The final levels chosen for each subject are shown in Table 2b. Because the target in each mixture had the same amplitude as the probe, and because the nontargets had current levels greater than zero, the loudness of the mixture was greater than that of the probe. Note that for one subject, CI5, the nontarget electrodes were set to -10% DR, meaning that each nontarget electrode was each stimulated at a level below the threshold for

TABLE 2

Part a shows the T and C levels for each electrode and subject, expressed in CUs. Part b shows the levels applied to the target and nontarget electrodes, expressed as a percentage of the DR for each electrode presented alone

	Subject				
	CI 1	CI 2	CI 3	CI 4	CI 5
a					
A					
T	151	202	211	164	194
C	179	238	245	228	222
B					
T	180	158	168	152	196
C	204	228	236	228	224
C					
T	163	154	200	164	186
C	203	234	234	224	224
D					
T	132	134	182	128	180
C	182	220	226	216	230
b					
Nontarget level (% DR)	32	25	27	25	-10
Target level (% DR)	51	48	60	70	70

that electrode alone. These nontarget electrodes were, however, almost certainly audible when presented together; the threshold for all four electrodes presented together at the same percent DR was -42% . It therefore appears that this subject showed a large amount of loudness summation. Our research software specified levels in CUs, which we converted to microamperes using the formula $\mu A = 10 \times (175^{CU/255})$. This formula, provided by Cochlear, was verified using a test implant and digital oscilloscope.

When loudness balancing was performed, the two stimuli to be balanced were presented in random order, and the subject indicated which one was the louder. The level of the variable stimulus was then adjusted on the next trial by an amount that was initially 10% DR. When the direction of change in the variable stimulus had reversed twice, the step size was reduced to 5% DR, and the procedure continued until 10 reversals had taken place. The current levels at the last eight reversals were then averaged. This procedure was performed four times, twice with the variable stimulus starting at a low level and twice at a level that was louder than that of the fixed stimulus. These four estimates were then averaged. When a multielectrode stimulus was varied, the levels applied to each electrode were varied by the same percentage of the DR, as measured for that electrode when presented alone.

Methods: experiment 1b

Experiment 1b investigated whether CI listeners can use rate, onset, or asynchrony cues in the absence of any additional current level boost applied to the target channel. The methods for experiment 1b were the same as for experiment 1a except that there was no level increment on the target channel; instead, all channels in the mixture were stimulated at the same percentage of their DR. Similarly, the probe was always presented at the same level as the corresponding channel in the mixture. There were no baseline conditions as, in the absence of onset, rate, or asynchrony cues, there was no a priori reason to expect the target channel to differ from any of the others. Two subjects, CI1 and CI2, were tested on all the conditions of experiments 1a and 1b. These subjects were selected for experiment 1b because they had shown a relatively large effect of onset delay in experiment 1a.

Procedure

In the main part of experiments 1a and 1b, each block of 20 trials consisted of a single condition with the target on electrode B on half of those trials and on electrode C on the other half. Subjects completed

one practice (10-trial block) and two blocks of each condition (40 trials) in turn before repeating the sequence of blocks in reverse order. The order was not formally counterbalanced but differed across subjects. This procedure was completed until there were at least 160 trials per condition. The results were analyzed both with the data averaged across the two targets and with the two target electrodes considered separately. Our discussion of the results will start with the former analysis.

Results: experiment 1a

Data averaged across the two target electrodes. Percent-correct scores averaged across the two target electrodes are shown for each subject in Figure 4a. Performance in the two baseline conditions (“base” and “BaseDrop”) did not differ significantly from each other, indicating that the final pulse of the stimulus did not have an effect on performance. In the following discussion, we compare the effects of each manipulation relative to the baseline condition and consider whether any such effects are greater than would be expected from random variation.

Performance in the Δ rate condition did not differ significantly from that in the baseline condition ($t=0.37$, $df=4$, $p=0.73$), suggesting that subjects could not use rate differences to “hear out” one channel in a mixture. This is consistent with our

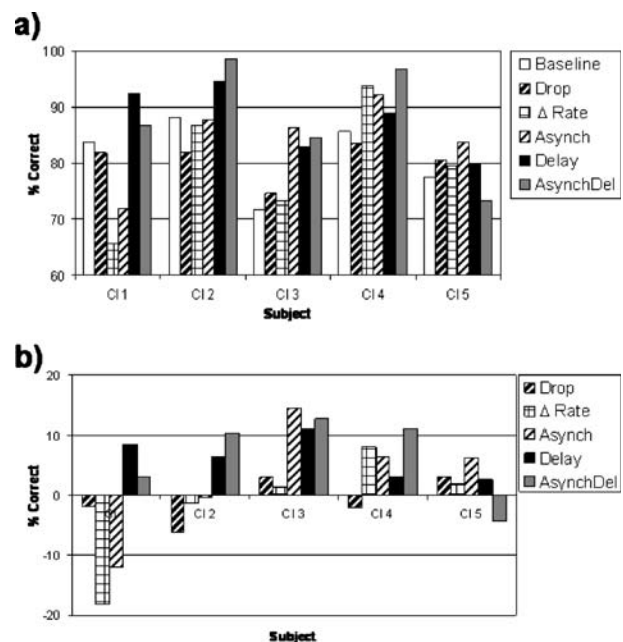


FIG. 4. (a) Performance for each subject and condition for experiment 1a. (b) The change in performance for various conditions relative to the baseline condition. In both cases, percent correct is averaged across trials when the target was on electrode B and when it was on electrode C.

previous finding, using simulations of CI hearing presented to NH listeners, that across-channel differences in rate do not provide a useful cue for concurrent sound segregation (Deeks and Carlyon 2004). Performance in the Asynch condition also did not differ significantly from baseline ($t=0.67$, $df=4$, $p=0.54$). However, there is some evidence that these two negative findings did not simply arise from a null effect or from random variation. Figure 4b shows performance in each condition with that in the baseline condition subtracted out; it can be seen that, when an asynchrony changed performance in a particular direction, a change in rate produced a change in the same direction. This relationship is reflected by a significant correlation between the change in performance *re* baseline in these two conditions ($r=0.84$, $df=4$, $p<0.05$). In principle, such a correlation between two difference scores could simply arise from the fact that the two differences have one score—that in the baseline condition—in common. Hence, even if scores were randomly distributed across conditions and subjects, then those subjects who happened to have a high score in the baseline condition would, on average, show a lower score when this baseline performance was subtracted out. However, the correlation between the two difference scores reported here remained significant when the correlation between each of these two difference scores and the baseline condition was partialled out (partial correlation = 0.97, $t=6.13$, $df=2$, $p<0.05$). Our interpretation is that an asynchrony can either enhance or impair performance, and that any effect of a rate difference is driven by the across-channel asynchrony that it produces (cf. Fig. 2). A caveat is that, in common with many CI experiments, we tested a fairly small number of subjects and that correlations obtained with

low numbers of observations should be treated with some caution.

The only condition to provide a significant improvement *re* the baseline condition was the Delay condition, in which there was a 200-ms onset delay on the target channel ($t=3.87$, $df=4$, $p=0.018$). This condition also differed significantly from the Base-Drop condition ($t=3.13$, $df=4$, $p=0.035$). However, it should be noted that this improvement was generally modest, ranging from 2 to 11 percentage points. The small size of the improvement is perhaps surprising, given the important role for onset delays that has been reported in several experiments with NH listeners (Darwin and Carlyon 1995). The remainder of the experiments reported here focus on the usefulness of such onset delays.

Data analyzed separately for the two target electrodes.

The percentage-correct scores are shown separately for each target and for each subject and condition in Table 3a. It can be seen that, for a given subject, performance could differ between the two types of target. When this happened, the difference in performance was usually consistent across conditions—for example, subject CI 3 always did better with the target on electrode C, whereas CI 5 usually did better with the target on electrode B. These differences in performance between the two target electrodes may have been due to a simple response bias, or to the “partial loudness” of the two targets not being equal. Differences in partial loudness could arise if partial masking produced by adjacent electrodes were not equal for the two targets.

Table 3a also shows that, although performance in the baseline condition was always below ceiling when averaged across the two electrodes (Fig. 4a), this was not always the case for each electrode when analyzed

TABLE 3

Results of experiments 1a (part a) and 1b (part b) shown separately for trials where the target was on channel B and on channel C

	CI 1		CI 2		CI 3		CI 4		CI 5	
	B	C	B	C	B	C	B	C	B	C
a										
Base	68.8	98.8	90.0	86.3	44.7	98.8	72.4	99.0	87.5	67.5
BaseDrop	65.0	98.8	90.0	73.8	50.6	98.8	67.1	100.0	86.3	75.0
Δ Rate	43.8	87.5	85.0	88.8	57.5	88.8	90.0	97.5	87.5	71.3
Asynch	47.5	96.3	94.1	81.2	76.3	96.3	85.6	98.9	85.1	82.4
Delay	85.7	98.8	91.1	97.8	65.9	100.0	78.8	98.8	79.0	81.0
AsynchDelay	73.8	100.0	99.0	98.0	70.0	98.9	96.8	96.8	77.5	68.8
b										
Δ Rate	2.5	78.8	56.3	58.8						
Asynch	2.5	77.5	56.3	53.8						
Delay	18.8	98.8	66.3	78.8						
AsynchDelay	13.8	92.5	76.5	96.5						

separately. This raises the possibility that the near-ceiling performance for some electrodes could have obscured differences between conditions. To test this idea, we repeated the analyses described in the previous section, taking into account only the data for the target electrode yielding the worst performance for each subject—electrode B for CI 1, CI 3, and C4, and electrode C for the other two subjects. The general pattern of results was the same as when the data were averaged across electrodes. Specifically, paired-sample t tests revealed no significant improvement *re* baseline in the Δ rate ($p=0.77$) and Asynch ($p=0.50$) conditions, but a significant improvement in the Delay condition ($p<0.01$). Performance in the AsynchDelay condition was also marginally ($p=0.05$) better than that in the baseline condition.

Results: experiment 1b

Data averaged across the two target electrodes. Figure 5 shows the results for the two subjects, CI 1 and CI 2, who took part in experiment 1b. Because, in this experiment, there was no consistent level cue to the target channel, we would expect that, in the absence of any additional cue, performance would be at chance (50%). It is therefore instructive to compare the scores obtained in each condition and for each subject to the chance level of 50%. Instances where scores differ significantly from chance are indicated by asterisks in the figure. A significant difference was defined as one where chance performance (50%) did not fall between the 95% confidence limits surrounding a given score.

The effects of the Δ rate and Asynch cues were consistent with those obtained for the same subjects in experiment 1a. For subject CI 1, performance in the Asynch and Δ rate conditions was worse than baseline in experiment 1 and significantly worse than chance in experiment 2. For subject CI 2, performance in these two conditions was similar to baseline

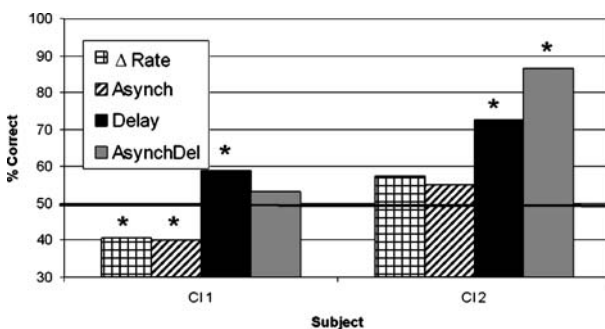


FIG. 5. Performance for two subjects in experiment 1b. Asterisks indicate conditions where performance differed significantly from chance (50%). Percent correct is averaged across trials when the target was on electrode B and when it was on electrode C.

in experiment 1 and close to chance in experiment 2. The fact that, for each subject, the effect of Δ rate and of an asynchrony was similar is consistent with our conclusion that the effect of a rate difference is driven by the asynchrony that it produces.

The effect of the 200-ms onset delay was also consistent with experiment 1a in that performance was significantly above chance for both subjects. No other differences were significant, except for subject CI 2 in the AsynchDelay condition. Performance for this condition and subject was also significantly above baseline in experiment 1a.

Data analyzed separately for the two target electrodes.

Performance analyzed for the two targets separately is shown in Table 3b. It can be seen that listener CI 1 shows a large difference in performance between the two targets, being below chance for electrode B and above chance for electrode C. This could be due to a response bias, or to the partial loudness for electrode C being higher, causing it to “pop out” of the mixture even when there was no additional cue to make it do so. The effect of this asymmetry would be to reduce the overall percentage correct when averaged across the two electrodes; an extreme asymmetry would limit performance to 50% in all conditions, even if the subject were sensitive to the cues introduced in particular conditions. Hence, the average scores shown in Figure 5 probably underestimate this subject’s underlying sensitivity. However, the asymmetry would strongly affect our interpretation of the pattern of results only if it varied markedly across conditions. To test this idea, we calculated a new measure in which we first defined a “hit” as the case where the subject correctly identified an interval when the target and probe were on electrode B, and a “false alarm” as the case when the target was on electrode C but the subject picked the interval when the probe was on electrode B. We then calculated a measure of asymmetry, or “bias,” in a way similar to that used to calculate the criterion “c” in a yes–no task: $\text{bias} = -0.5[z(H) + z(FA)]$ (Macmillan and Creelman 1991). The absence of any bias would yield a score of zero. The resulting values were similar across conditions for both subjects: for CI 1 they varied between 1.24 and 1.48 across conditions, and for CI 2 they ranged from -0.03 to $+0.52$.

EXPERIMENT 2: INVESTIGATING THE ROLE OF ONSET CUES WITH NH LISTENERS

Rationale and conditions

Although experiment 1 showed that CI users can exploit onset differences in concurrent sound segregation, the small size of the advantage conveyed by onset delays is perhaps surprising, given that onset

cues provide one of the strongest cues for segregation by NH listeners (Darwin and Carlyon 1995). Experiment 2 probed the use of onset difference by NH listeners using a task that was roughly analogous to that performed by the CI listeners of experiment 1. By manipulating various stimulus parameters, and observing the effect on performance, we aimed to identify the aspects of peripheral stimulation that strongly influence whether listeners can or cannot exploit onset differences.

The stimuli used in two example trials of the “Pulse” condition are shown schematically in Figure 6. As in experiment 1b, each half of a two-interval forced-choice trial consisted of a “probe” followed by a “mixture.” The mixture consisted of harmonics 5–40 of a 100-Hz F0, with a duration of 400 ms, and where the onsets of harmonics 13–17 or 26–30 were delayed by 200 ms. These “target” harmonics are all unresolved by the peripheral auditory system, and, because they were summed in sine phase, auditory filters centered on them had outputs that resembled filtered 100-Hz pulse trains. We have previously reported analogous results obtained with pulse trains produced from unresolved harmonics and presented to NH listeners, and electric pulse trains presented to CI users (McKay and Carlyon 1999; Carlyon et al. 2002;

van Wieringen et al. 2003). The only difference in the paradigm, compared to experiment 1b, was that the target band of harmonics was not constrained to be the same on the two halves of each trial. The choice of which band was delayed was taken independently for each half of the trial, and the probe was selected to match the delayed band in only one half.

The pulsatile stimuli used in experiment 1 would have produced coherent amplitude modulation (AM) in the responses of neurons tuned to each electrode. Such coherent AM may have caused the responses of those neurons to have been grouped by the central auditory system, and this grouping may have partially counteracted any effect of onset asynchrony. A similar phenomenon may have occurred in the Pulse condition of experiment 2. To reduce this AM we included a “Noise” condition, in which the F0 of the harmonic complex was reduced to 2.5 Hz and the harmonics were summed in random phase. The spectral regions of the complex and of the two target bands were the same as in the Pulse condition, but the waveform at the output of each auditory filter resembled a band of noise instead of a filtered pulse train.

Many of the previous experiments demonstrating a strong role for onset asynchronies in concurrent sound segregation have used low-frequency components of harmonic complexes (Darwin and Sutherland 1984; Darwin and Ciocca 1992; Carlyon 1994; Hukin and Darwin 1995). Such harmonics, being well-resolved by the peripheral auditory system, excite discrete sets of AN fibers. In contrast, the broad spread of excitation often observed with electrical stimulation may have caused the excitation produced by the four electrodes in experiment 1 to have overlapped. To test the role of such overlap, we included the “PulseGap” condition; this was identical to the “Pulse” condition except that harmonics adjacent to the target bands (numbers 10–12, 18–19, 24–25, and 31–34) were absent.

Finally, four conditions were included in which the complex consisted of four well-resolved partials, and where the targets and probes were either the 2nd- or 3rd-highest partial. These conditions were included to test the influence of spectral region and of harmonicity on the ability of listeners to use onset differences in concurrent sound segregation. In the “LowHarm” condition, the partial frequencies were 400, 800, 1,200, and 1,600 Hz. The “LowLog” condition was similar except that the complex was inharmonic, with adjacent partials separated by a factor of 1.5, so that they were equally spaced on a log scale; their frequencies were 533, 800, 1,200, and 1,800 Hz. The HiLog condition was similar except that the partial frequencies were a factor of 2.63 higher than in the LowLog condition, and so were 1,401, 2,102, 3,153, and 4,370 Hz. Finally, as an additional test of

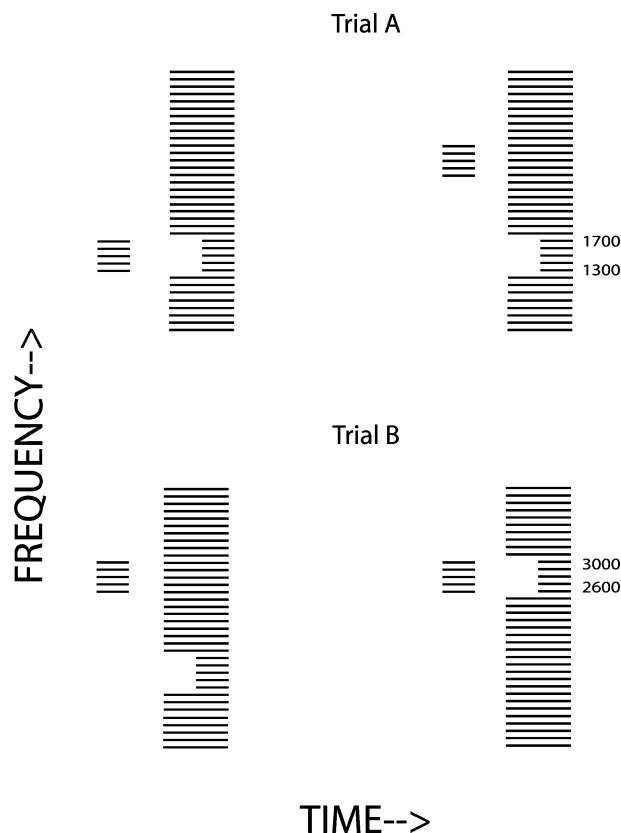


FIG. 6. Schematic spectrogram of the stimuli in two trials in the Pulse condition of experiment 2. The signal is in interval 1 of trial A and in interval 2 of trial B.

the role of spectral overlap, another high-frequency inharmonic condition was included, in which the ratio between adjacent partials was reduced from 1.5 to 1.236. In this “HiNarrow” condition, the partial frequencies were 1,700, 2,102, 2,600, and 3,215 Hz.

Stimulus generation

The probes and mixtures were generated digitally and played out at a sampling rate of 20,000 Hz from one DAC of a CED1401 laboratory interface. They were then lowpass filtered (Kemo VBF25.01; cut-off=8,600 Hz, attenuation rate=100 dB/octave), attenuated (TDT PA2), and passed to one input of a headphone amplifier. A continuous pink noise, produced by passively filtering a white noise (TDT WG1), was attenuated and mixed with the other stimuli by feeding it to another input of the headphone amplifier. The spectrum level of the pink noise at 1 kHz was 10-dB sound pressure level (SPL). The stimuli were presented at a level of 50 dB SPL per component, except in the Noise condition where the overall level was the same as in the Pulse and PulseGap conditions. This level was 65.6 dB SPL during those times when all components were present in the mixture. All components were turned on and off with 10-ms raised-cosine ramps, except in the LowHarm and LowLog conditions where the ramp duration was increased to 26.3 ms in order to minimize the effects of spectral splatter. All frequency components in any given stimulus had the same amplitude, and those in each probe had the same amplitude per component as in the mixtures used in the corresponding condition.

Procedure

Five NH subjects took part and were tested individually in a double-walled sound-insulating booth. Each listened through one earpiece of a Sennheiser HD250 headset. Each experimental run consisted of 25 trials of a single condition. No feedback was provided. Subjects were told to identify the interval in which the probe was present in the mixture. Conditions were presented in a counterbalanced order until each subject had completed 200 runs per condition. Results were expressed in terms of the percentage of trials in which the subject correctly identified the interval in which the probe corresponded to the delayed component(s).

Results

Pulse, Noise, and PulseGap conditions. Figure 7a shows performance in the Pulse, Noise, and PulseGap conditions of experiment 2, with mean performance

plotted in the right-most cluster of bars. Performance in the “Pulse” condition was modest, with a mean score (black bar) of 73%. Hence, even in NH listeners, the presence of a 200-ms onset delay is not always sufficient to produce good sound segregation. The results obtained in the Noise condition are shown by the white bar. On average, performance was similar to that in the Pulse condition, although subject NH3 showed performance that was substantially better. The similar average performance in the two conditions suggests that the relatively low level of performance in the Pulse condition was not due to the pulsatile nature of the auditory-nerve response to the target bands. In contrast, introducing a spectral gap between the target and nontarget bands (PulseGap condition, gray bars) produced a substantial improvement in performance that is consistent across subjects. Figure 7b shows that performance was also good in the remaining conditions, where the stimuli consisted of four concurrent pure tones.

The above trends were supported by a one-way ANOVA, based on all seven conditions, followed by a series of planned comparisons. The ANOVA revealed a highly significant main effect of condition [$F(6,24) = 4.77$; $p < 0.01$; Huynh–Feldt sphericity correction]. Planned comparisons showed no difference between the Noise and Pulse conditions but that, compared to the Pulse condition, there was a

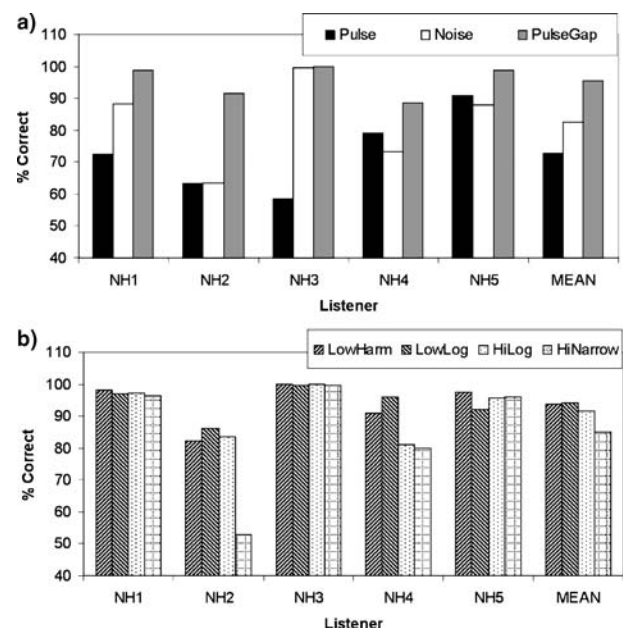


FIG. 7. The five left-most clusters of bars in (a) show performance, measured in percent correct, for five NH listeners in the Pulse, PulseGap, and Noise conditions of experiment 2. The right-most cluster of bars is similar to (a) but shows performance averaged across listeners. (b) is similar to (a) but shows performance in those conditions where the targets consisted of single frequency components.

significant performance elevation in the PulseGap condition.

Conditions involving four resolved components. Performance for those conditions involving four tonal components is shown in Figure 7b. Generally speaking, performance in this latter group of conditions was close to ceiling for at least some subjects, a finding which may have obscured any true differences between them. Planned comparisons revealed that performance in the LowHarm and LowLog conditions was superior to that in the Pulse condition. However, performance in these two conditions did not differ significantly from that in the other conditions involving four tonal components (HiHarm and HiLog).

The main conclusion to be drawn from experiment 2 is that a major factor limiting the usefulness of onset cues is the spectral separation between the target and nontarget components, and that other factors—such as overall frequency region or the pulsatile nature of AN responses—do not appear to impose such a major limitation.

DISCUSSION

Use of onset-time cues by NH and CI listeners

The results of experiment 1 showed that CI users obtained a small but modest improvement in concurrent sound segregation when stimulation applied to the target electrode started 200 ms later than that applied to the others. Experiment 2 showed that a similarly modest improvement could be obtained with NH listeners and acoustic stimulation, and that performance could be improved markedly by introducing a spectral gap between the target and nontarget bands. The results of the latter experiment suggest that overlap between the internal excitation patterns of target and nontarget stimuli can reduce the effectiveness of onset cues. One way in which this could happen is that, to the extent that neurons tuned to the target stimulus will already be stimulated by nontarget stimuli, the increase in their response produced by the target onset will be reduced. This is illustrated in Figure 8, which shows the excitation patterns in the Pulse condition calculated according to the method described by Moore et al. (1997). The solid line shows the case where all harmonics are present, whereas the dashed line shows the excitation pattern when the upper target band (2,600–3,000 Hz) is absent. It can be seen that the maximum change of excitation produced by the target is about 9 dB, which, although easily detectable (Green 1988), is substantially less than the 22 dB occurring in the PulseGap condition (not shown). If there were substantial overlap between the excitation patterns

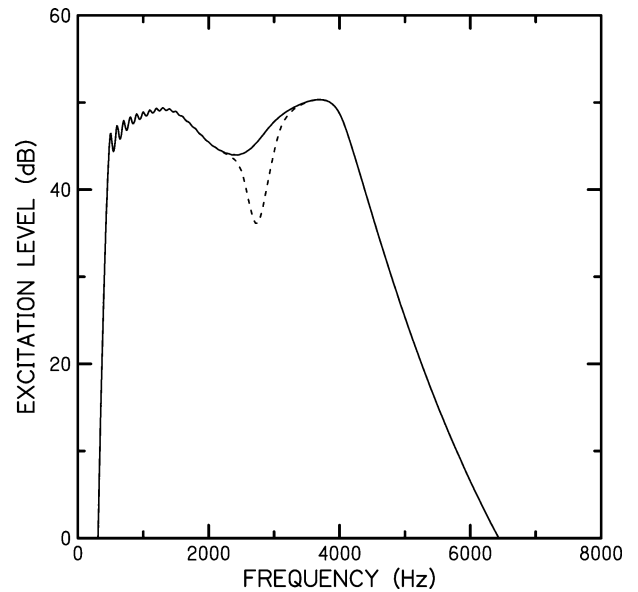


FIG. 8. Excitation patterns, derived from the procedure described by Moore et al. (1997), for the Pulse condition of experiment 2. The solid line shows the case where all components are present. The dotted line shows the case where harmonics 26–30 have not yet been turned on; there is a 9-dB dip in the excitation pattern relative to the solid line. A similar but larger dip occurred when the lower target harmonics (13–17) were absent.

produced by the four electrodes in experiment 1, then a similar phenomenon could account for the modest size of the improvements observed in that experiment. Measures of spread-of-excitation, obtained by forward-masking patterns, in CI listeners vary across subjects, but there is evidence of significant amounts of masking between electrodes separated by the 3 mm used in the present study (Chatterjee et al. 2006; Kwon and van den Honert 2006). We feel that this is the most likely explanation for the modest size of the advantage obtained from onset cues by CI users. We should note, however, that comparisons of the spread-of-excitation in NH and CI are hampered by a number of factors. These include the difficulty in equating, across electric and acoustic hearing, the ordinate to be used when plotting excitation patterns (Kwon and van den Honert 2006), and the level-dependent asymmetrical spread of masking in NH, which does not occur consistently in CIs.

Our discussion so far has focused on the use of onset cues to segregate a target band of stimulation that starts later than nontarget bands. In real life, onset differences will also be useful because the sound one wants to listen to may start before other, interfering, sounds. This may allow two benefits.

First, it could allow the listener an uninterrupted “glimpse” of the target sound. Nelson et al (2003) measured speech perception in the presence of speech-shaped noise that was either steady or modulated at rates of between 1 and 32 Hz. NH listeners

performed substantially better at all modulation rates compared to steady noise. However, neither CI users nor NH listeners listening to a “vocoder” implant simulation showed a consistent advantage, although the CI users did show some small improvement at the lowest and highest modulation rates tested. Nelson et al. suggested that the absence of a consistent advantage could be due to the CIs not preserving sufficient spectral detail for the listeners to distinguish between the modulated noise and the speech. Evidence in support of this interpretation comes from a recent study by Fu and Nogaki (2005). In one set of conditions, they measured speech reception thresholds (SRTs) in steady and modulated maskers for NH subjects listening to vocode simulations. By varying the number of channels and amount of spectral smearing in the simulations, they showed that the reduction in SRT produced by modulating a masker depended strongly on the amount of spectral resolution preserved by different noise-vocoder simulations. The results of Nelson et al. and of Fu and Nogaki suggest that, for the perception of sentences in modulated noise, as for the paradigm tested here, the usefulness of timing differences between the target and the masker depends strongly on them exciting separate neural populations.

Second, when the target starts before other components, the listener might obtain a better estimate of which neurons’ responses “belong” to the target sound, so that this neural excitation can be subsequently subtracted from the mixture (Dannenbring and Bregman 1978). However, at least for a pitch perception task, this strategy does not appear to be useful for NH listeners when the target and interfering sounds excite the same neural population (Carlyon 1996b). So, even under these circumstances, the use of onset cues may depend strongly on the extent to which the target and nontarget sounds excite discrete populations of AN fibers.

Use of temporal pitch differences by CI users

Experiment 1 showed that CI users were not consistently better at sound segregation when the target band was stimulated at a different rate to the others, compared to when they were stimulated at the same rate. This result is consistent with the findings of Deeks and Carlyon (2004), who modified the popular “noise vocoder” simulation of CIs so that the envelope in each frequency band modulated a filtered acoustic pulse train, whose rate was such that all of its harmonics were unresolved by the peripheral auditory system. In one condition, a speech signal and a time-reversed speech masker were split into nonoverlapping “analysis” bands, and the envelope in each band modulated a pulse train that was

filtered in the same way as the analysis band. Performance was not consistently better when the pulse trains conveying the target speech had a different rate from those conveying the masker, compared to when all pulse trains had the same rate (instead, performance depended primarily on the pulse rate used for the target speech, with a rate of 140 pps being better than an 80-pps rate). These results are also consistent with Darwin’s (1992) finding that F0 differences between one formant of a speech sound and the rest did not influence the integration of that formant into the sound’s phonetic identity when its harmonics were unresolved by the peripheral auditory system. Overall, it seems that differences in F0 have a strong effect on *concurrent* sound segregation only when conveyed by resolved harmonics in NH. This could be simply because such harmonics produce a stronger pitch percept than other sounds (Hoekstra 1979; Houtsma and Smurzynski 1990; Shackleton and Carlyon 1994; Gockel et al. 2004), or perhaps because the auditory system exploits a deviation from spectral regularity, which does not occur when F0 differences are encoded purely by the temporal response of the AN (Roberts 2005).

Effects of across-channel asynchrony

Experiment 1 showed that a 5-ms asynchrony between the target and non target channels could sometimes help and sometimes hinder performance. This finding was consistent across experiments 1a and 1b, and the effect across listeners was similar to that produced by a rate difference. We concluded that the effect of a rate difference was driven by the asynchrony that it produced, and it appears that the effects of an asynchrony are reliable but idiosyncratic. The results may be due to two competing factors, both of which can be illustrated with reference to the model described by (McKay et al. 2001; McKay et al. 2003), and which calculates the specific loudness within groups of AN fibers tuned to similar frequencies. An asynchrony could increase the specific loudness of a target band by reducing refractory effects produced by neighboring electrodes stimulating fibers tuned to the target. Conversely, because the excitation produced by adjacent pulses is summed over a 2-ms window, then the specific loudness produced by target pulses will be augmented by that of non target pulses, and this effect will be reduced by a 5-ms asynchrony.

Implications for the development of CIs

The results and discussion presented here have some implications for the development of future CIs, and for which developments are most likely to improve

sound segregation by implant users. One aim of much contemporary CI research is to improve the encoding of F0. Much of this effort has been directed towards modifying the existing way of encoding F0 so as to enhance the perception of the F0-rate modulations imposed on the pulse trains presented to each channel. Modest improvements in pitch perception have been achieved by increasing modulation depth, aligning the modulations across channels, and changing the shape of the modulator (McDermott et al. 1992; Green et al. 2004; Laneau et al. 2006). The present results suggest that this approach will not improve the segregation of *simultaneous* portions of competing voices. It may, however, improve the enjoyment of music and the perception of prosody when the stimulus is dominated by a single sound source. It is also possible that such an approach could aid sequential sound segregation: measures of auditory streaming by NH listeners show that it can be enhanced by differences in F0, even when these are conveyed by “purely temporal” aspects of the peripheral response (Vliegen et al. 1999; Vliegen and Oxenham 1999; Grimault et al. 2000). An alternative would be to try to “reproduce” the encoding of resolved harmonics in a CI. However, this may involve aspects of the NH neural response that would be difficult to replicate in a CI. These may include a close match between place and rate of stimulation (Oxenham et al. 2004; Moore and Carlyon 2005), a reliable phase transition around the peak of the traveling wave (Kim et al. 1980; Shamma 1985; Loeb 2005; Moore and Carlyon 2005), and the ability to convey periodicities of up to a few thousand Hertz, despite evidence that most CI users do not exploit temporal cues to pitch at rates above about 300 Hz (Shannon 1983; Townshend et al. 1987; McKay et al. 2000).

A different approach to improving concurrent sound segregation has been, instead of concentrating on F0 encoding, to modify either the electrode design or stimulation mode in an attempt to improve channel selectivity (Bierer and Middlebrooks 2004; Snyder et al. 2004). The results presented here suggest that a successful application of that approach would produce substantial improvements in concurrent sound segregation. Similar advantages would occur for sequential sound segregation, because auditory streaming is often greatest when the competing sounds excite separate populations of AN fibers (Hartmann and Johnson 1991; Vliegen et al. 1999; Grimault et al. 2000). However, advantages of improved channel selectivity have not yet been comprehensively demonstrated in human patients. For example, although bipolar and tripolar stimulation can produce more restricted excitation than monopolar stimulation in animal physiological recordings (Bierer and Middlebrooks 2004; Snyder

et al. 2004), such advantages have not yet been observed using behavioral measures in humans. Kwon and van den Honert (2006) reported that the widths of forward-masked excitation patterns do not differ consistently between equally loud maskers presented in monopolar compared to bipolar mode. They suggested that although the spread of current may be narrower for bipolar than for monopolar maskers of the same level, the increase in level needed to make the bipolar masker equally loud may increase the spread of current that it produces. Hence, the challenge may be to design a form of stimulation that, in neurons close to the active electrode, causes a firing rate that is sufficiently high to produce a loud percept without requiring the level to be so high as to produce a wide current spread.

SUMMARY

CI listeners could not use across-channel differences in pulse rate (100 vs 71 pps) to perceptually segregate stimulation on one “target” channel of a four-channel stimulus. CI listeners could use differences in onset time, although the benefit conferred by a quite large onset time differences of 200 ms was quite modest. NH listeners also obtained a modest level of performance when required to segregate a delayed band of unresolved harmonics from a complex tone consisting of harmonics 5 to 40 of a 100-Hz F0. The ability of NH listeners to exploit onset differences improved markedly when a spectral gap was introduced between the target and nontarget harmonics. Greatly superior performance was observed in a range of conditions in which the complex was replaced by four well-separated components, and where subjects had to segregate one of these from the rest. It is suggested that the use of onset time differences for concurrent sound segregation depends strongly on the target and nontarget stimuli exciting discrete populations of AN fibers, and can be impaired in CI users by a spread of current along the cochlea.

ACKNOWLEDGMENT

This research was supported by a grant from the Leverhulme Trust to the first author. We thank Monita Chatterjee for helpful comments on a previous version of this article.

REFERENCES

- BAUMANN U, NOBBE A. Pulse rate discrimination with deeply inserted electrode arrays. *Hear. Res.* 196:49–57, 2004.

- BENCH J, BAMFORD J. Speech-Hearing Tests and the Spoken Language of Hearing-Impaired Children. London, Academic, 1979.
- BIERER JA, MIDDLEBROOKS JC. Cortical responses to cochlear implant stimulation: channel interactions. *J. Assoc. Res. Otolaryngol.* 5:32-48, 2004.
- BREGMAN AS. Auditory Scene Analysis. Cambridge, MA, M.I.T. Press, 1990.
- BREGMAN AS, PINKER S. Auditory streaming and the building of timbre. *Can. J. Psychol.* 32:19-31, 1978.
- CARLYON RP. Detecting mistuning in the presence of synchronous and asynchronous interfering sounds. *J. Acoust. Soc. Am.* 95:2622-2630, 1994.
- CARLYON RP. Encoding the fundamental frequency of a complex tone in the presence of a spectrally overlapping masker. *J. Acoust. Soc. Am.* 99:517-524, 1996a.
- CARLYON RP. Masker asynchrony impairs the fundamental-frequency discrimination of unresolved harmonics. *J. Acoust. Soc. Am.* 99:525-533, 1996b.
- CARLYON RP, GOCKEL H. Effects of harmonicity and regularity on the perception of sound sources. In: Yost WA (ed) *Springer Handbook of Auditory Research: Auditory Perception of Sound Sources*. Berlin, Springer-Verlag (in press).
- CARLYON RP, VAN WIERINGEN A, LONG CJ, DEEKS JM, WOUTERS J. Temporal pitch mechanisms in acoustic and electric hearing. *J. Acoust. Soc. Am.* 112:621-633, 2002.
- CHATTERJEE M, GALVIN JJ, FU QJ, SHANNON RV. Effects of stimulation mode, level and location on forward-masked excitation patterns in cochlear implant patients. *J. Assoc. Res. Otolaryngol.* 7:15-25, 2006.
- DANNENBRING GL, BREGMAN AS. Streaming vs. fusion of sinusoidal components of complex tones. *Percept. Psychophys.* 24:369-376, 1978.
- DARWIN CJ. Perceptual grouping of speech components differing in fundamental frequency and onset time. *Q. J. Exp. Psychol.* 33A:185-287, 1981.
- DARWIN CJ. Listening to two things at once. In: Schouten B (ed) *Audition, Speech, and Language*. Berlin, Mouton-De Gruyter, pp. 133-148, 1992.
- DARWIN CJ, CARLYON RP. Auditory grouping. In: Moore BCJ (ed) *Hearing*. Orlando, FL, Academic, pp. 387-424, 1995.
- DARWIN CJ, CIOCCA V. Grouping in pitch perception: effects of onset asynchrony and ear of presentation of a mistuned component. *J. Acoust. Soc. Am.* 91:3381-3390, 1992.
- DARWIN CJ, GARDNER RB. Mistuning a harmonic of a vowel: grouping and phase effects on vowel quality. *J. Acoust. Soc. Am.* 79:838-845, 1986.
- DARWIN CJ, SUTHERLAND NS. Grouping frequency components of vowels: when is a harmonic not a harmonic? *Q. J. Exp. Psychol.* 36A:193-208, 1984.
- DEEKS JM, CARLYON RP. Simulations of cochlear implant hearing using filtered harmonic complexes: implications for concurrent sound segregation. *J. Acoust. Soc. Am.* 115:1736-1746, 2004.
- FU QJ, NOGAKI G. Noise susceptibility of cochlear implant users: the role of spectral resolution and smearing. *J. Assoc. Res. Otolaryngol.* 6:19-27, 2005.
- GOCKEL H, CARLYON RP, PLACK CJ. Across-frequency interference effects in fundamental frequency discrimination: questioning evidence for two pitch mechanisms. *J. Acoust. Soc. Am.* 116:1092-1104, 2004.
- GREEN DM. Profile Analysis. New York, Oxford University Press, 1988.
- GREEN T, FAULKNER A, ROSEN S. Enhancing temporal cues to voice pitch in continuous interleaved sampling cochlear implants. *J. Acoust. Soc. Am.* 116:2298-2310, 2004.
- GRIMAULT N, MICHEYL C, CARLYON RP, ARTHAUD P, COLLET L. Influence of peripheral resolvability on the perceptual segregation of harmonic complex tones differing in fundamental frequency. *J. Acoust. Soc. Am.* 108:263-271, 2000.
- HARTMANN WM, JOHNSON D. Stream segregation and peripheral channeling. *Music Percept.* 9:155-184, 1991.
- HARTMANN WM, McADAMS S, SMITH BK. Hearing a mistuned harmonic in an otherwise periodic complex tone. *J. Acoust. Soc. Am.* 88:1712-1724, 1990.
- HOEKSTRA A. Frequency discrimination and frequency analysis in hearing. Ph.D. Thesis, Institute of Audiology, University Hospital, Groningen, Netherlands, 1979.
- HOUTSMA AJM, SMURZYNSKI J. J.F. Schouten revisited: pitch of complex tones having many high-order harmonics. *J. Acoust. Soc. Am.* 87:304-310, 1990.
- HUKIN RW, DARWIN CJ. Comparison of the effect of onset asynchrony on auditory grouping in pitch matching and vowel identification. *Percept. Psychophys.* 57:191-196, 1995.
- KETTEN DR, VANNIER MW, SKINNER MW, GATES GA, WANG G, NEELY JG. In vivo measures of cochlear length and insertion depth of nucleus cochlear implant electrode arrays. *Ann. Otol. Rhinol. Laryngol.* 107:1-16, 1998.
- KIM DO, MOLNAR CE, MATTHEWS JW. Cochlear mechanics: nonlinear behavior in two-tone responses as reflected in cochlear-nerve fiber responses and in ear-canal sound pressure. *J. Acoust. Soc. Am.* 67:1704-1721, 1980.
- KWON BJ, VAN DEN HONERT C. Effect of electrode configuration on psychophysical forward masking in cochlear implant listeners. 119:2994-3002, 2006.
- LANEAU J, WOUTERS J, MOONEN M. Improved music perception with explicit pitch coding in cochlear implants. *Audiology and Neurotology* 11:38-52, 2006.
- LEVITT H. Transformed up-down methods in psychophysics. *J. Acoust. Soc. Am.* 49:467-477, 1971.
- LOEB GE. Are cochlear implant patients suffering from perceptual dissonance? *Ear Hear.* 26:435-450, 2005.
- MACMILLAN NA, CREELMAN CD. Detection Theory: A User's Guide. Cambridge, Cambridge University Press, 1991.
- McDERMOTT HJ, McKAY CM, VANDALI AE. A new portable sound processor for the University-of-Melbourne nucleus limited multielectrode cochlear implant. *J. Acoust. Soc. Am.* 91:3367-3371, 1992.
- McKAY CM, CARLYON RP. Dual temporal pitch percepts from acoustic and electric amplitude-modulated pulse trains. *J. Acoust. Soc. Am.* 105:347-357, 1999.
- McKAY CM, HENSHALL KR, FARREL RJ, McDERMOTT HJ. A practical method of predicting the loudness of complex electrical stimuli. *J. Acoust. Soc. Am.* 113:2054-2063, 2003.
- McKAY CM, McDERMOTT HJ. Loudness perception with pulsatile electrical stimulation: the effect of interpulse intervals. *J. Acoust. Soc. Am.* 104:1061-1074, 1998.
- McKAY CM, McDERMOTT HJ, CARLYON RP. Place and temporal cues in pitch perception: are they truly independent? *Acoust. Res. Lett. Online* (<http://scitation.aip.org/arlo/>) 1:25-30, 2000.
- McKAY CM, RENNINE MD, McDERMOTT HJ. Loudness summation for pulsatile electrical stimulation of the cochlea: Effects of rate, electrode separation, level, and mode of stimulation. *J. Acoust. Soc. Am.* 110:1514-1526, 2001.
- MOORE BCJ. An Introduction to the Psychology of Hearing, 2nd edn. London, Academic, 1982.
- MOORE BCJ, CARLYON RP. Perception of pitch by people with cochlear hearing loss and by cochlear implant users. In: Plack CJ, Oxenham AJ (eds) *Springer Handbook of Auditory Research: Pitch Perception*. Berlin, Springer-Verlag, pp. 234-277, 2005.
- MOORE BCJ, GLASBERG BR, BAER T. A model for the prediction of

- thresholds, loudness and partial loudness. *J. Audio Eng. Soc.* 45:224–240, 1997.
- NELSON PB, JIN SH, CARNEY AE, NELSON DA. Understanding speech in modulated interference: cochlear implant users and normal-hearing listeners. *J. Acoust. Soc. Am.* 113:961–968, 2003.
- OXENHAM AJ, BERNSTEIN JGW, PENAGOS H. Correct tonotopic representation is necessary for complex pitch perception. *Proc. Natl. Acad. Sci. USA* 101:1421–1425, 2004.
- ROBERTS B. Spectral pattern, grouping, and the pitches of complex tones and their components. *Acta Acustica United Acustica* 91:945–957, 2005.
- SHACKLETON TM, CARLYON RP. The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination. *J. Acoust. Soc. Am.* 95:3529–3540, 1994.
- SHAMMA S. Speech processing in the auditory system: II. Lateral inhibition and the central processing of speech evoked activity in the auditory nerve. *J. Acoust. Soc. Am.* 78:1622–1632, 1985.
- SHANNON RV. Multichannel electrical stimulation of the auditory nerve in man. I. Basic psychophysics. *Hear. Res.* 11:157–189, 1983.
- SKINNER MW, CLARK GM, WHITFORD LA, ET AL. Evaluation of a new spectral peak coding strategy for the Nucleus 22 channel cochlear implant system. *Am. J. Otol.* 15:15–27, 1994.
- SNYDER RL, BIERER JA, MIDDLEBROOKS JC. Topographic spread of inferior colliculus activation in response to acoustic and intracochlear electric stimulation. *J. Assoc. Res. Otolaryngol.* 5:305–322, 2004.
- STICKNEY GS, ZENG FG, LITOVSKY R, ASSMANN P. Cochlear implant speech recognition with speech maskers. *J. Acoust. Soc. Am.* 116:1081–1091, 2004.
- TONG YC, CLARK GM. Absolute identification of electric pulse rates and electrode positions by cochlear implant listeners. *J. Acoust. Soc. Am.* 74:73–80, 1985.
- TOWNSHEND B, COTTER N, COMPERNOLLE DV, WHITE RL. Pitch perception by cochlear implant subjects. *J. Acoust. Soc. Am.* 82:106–115, 1987.
- VAN WIERINGEN A, CARLYON RP, LONG CJ, WOUTERS J. Pitch of amplitude-modulated irregular-rate stimuli in electric and acoustic hearing. *J. Acoust. Soc. Am.* 114:1516–1528, 2003.
- VLIEGEN J, OXENHAM AJ. Sequential stream segregation in the absence of spectral cues. *J. Acoust. Soc. Am.* 105:339–346, 1999.
- VLIEGEN J, MOORE BCJ, OXENHAM AJ. The role of spectral and periodicity cues in auditory stream segregation, measured using a temporal discrimination task. *J. Acoust. Soc. Am.* 106:938–945, 1999.
- ZENG F-G. Temporal pitch in electric hearing. *Hear. Res.* 174:101–106, 2002.