

# Simulations of cochlear implant hearing using filtered harmonic complexes: Implications for concurrent sound segregation

John M. Deeks and Robert P. Carlyon

*MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 2EF, United Kingdom*

(Received 5 September 2003; revised 14 January 2004; accepted 23 January 2004)

Two experiments used simulations of cochlear implant hearing to investigate the use of temporal codes in speech segregation. Sentences were filtered into six bands, and their envelopes used to modulate filtered alternating-phase harmonic complexes with rates of 80 or 140 pps. Experiment 1 showed that identification of single sentences was better for the higher rate. In experiment 2, maskers (time-reversed concatenated sentences) were scaled by  $-9$  dB relative to a target sentence, which was added with an offset of 1.2 s. When the target and masker were each processed on all six channels, and then summed, processing the masker on a different rate to the target improved performance only when the target rate was 140 pps. When the target sentence was processed on the odd-numbered channels and the masker on the even-numbered channels, or vice versa, performance was worse overall, but showed similar effects of pulse rate. The results, combined with recent psychophysical evidence, suggest that differences in pulse rate are unlikely to prove useful for concurrent sound segregation. © 2004 Acoustical Society of America. [DOI: 10.1121/1.1675814]

PACS numbers: 43.66.Ts, 43.71.Ky [GK]

Pages: 1736–1746

## I. INTRODUCTION

Normally hearing listeners can use differences in fundamental frequency ( $\Delta F_0$ s) between concurrent speech sounds to segregate them perceptually (Brokx and Nootboom, 1982; Scheffers, 1983; Assmann and Summerfield, 1990; Culling and Darwin, 1993). For example, introducing a  $\Delta F_0$  between the members of a pair of concurrent vowels can lead to a significant improvement in the ability of listeners to identify both members of the pair (Scheffers, 1983). Subjectively, the improvement is often accompanied by an impression of two voices producing different vowels on different pitches (Assmann and Summerfield, 1990). Assmann and Summerfield (1990) presented pairs of concurrent vowels with  $\Delta F_0$ s between them to normal-hearing listeners. They showed that the pattern of vowel identification could be well fitted by a model that combined the effects of filtering in the auditory periphery with an analysis of the temporal pattern of activity predicted to occur in the auditory nerve. A similar spectro-temporal approach has been adopted by Meddis and Hewitt (1992) and by Brown and Cooke (1994). Darwin (1992) has shown that frequency components that are resolved by the peripheral auditory system contribute most to the segregation process. In the normal auditory system, the individual frequencies of these resolved harmonics are encoded both by their place of excitation, and by the temporal responses of subsets of auditory-nerve fibers tuned to those frequencies.

Encoding of  $F_0$  in a cochlear implant differs from that found in normal acoustic hearing (Moore and Carlyon, 2004). The incoming signal is passed through a bank of filters that are too wide to resolve individual harmonics, and, in most modern strategies, used to modulate a pulse train having a fixed rate. Hence, there is little opportunity for subsets of stimulated fibers to encode the temporal structure of an individual, resolved, harmonic, as occurs in a normal ear. However, for a periodic sound presented in quiet,  $F_0$  can be

represented as temporal fluctuations in the envelope resulting from harmonics interacting within the passband of each channel. Geurts and Wouters (2001) measured just-noticeable differences in synthetic vowel  $F_0$ , by implant users whose device implemented the continuous interleaved sampling (CIS) processing strategy (Wilson *et al.*, 1991). For a vowel with an  $F_0$  of 150 Hz, their four subjects could hear differences of between 4.0% and 13.3%. However, their task involved the sequential presentation of  $\Delta F_0$ , and it is likely that different processes are involved in the use of  $\Delta F_0$  to segregate concurrent voices. Indeed, unlike normally hearing listeners, implant users are unable to exploit differences between the gender of a target speaker and that of an interfering speaker, consistent with them being unable to use  $F_0$  differences for concurrent sound segregation. For these listeners, competing speech can impair performance even at a favorable signal-to-noise ratio of  $+16$  dB (Nelson and Jin, 2002).

Further insight into the difficulties implant users face when attempting to use pitch cues to separate competing sounds comes from acoustic simulations of cochlear implant speech processors (Shannon *et al.*, 1995). Typically, these extract the temporal envelope in each of several frequency regions, and use this envelope to modulate a band of noise or sinusoidal carrier. When listening to such simulations, normally hearing subjects are also able to detect  $F_0$  differences of a few percent between sequentially presented sounds, provided that these have a reasonably low  $F_0$  and are not presented in a reverberant environment (Qin and Oxenham, 2003a). However, also consistent with the cochlear implant literature, they are more susceptible to interfering noise than when they listen to unprocessed speech (Qin and Oxenham, 2003b).

One reason for this deficit may lie in the inability of the auditory system to use purely temporal cues to extract the  $F_0$ s of two periodic stimuli that excite the same region of the

cochlea. This evidence comes from studies in which mixtures of two periodic pulse trains were either applied to a single cochlear implant electrode, or presented to normally hearing listeners after filtering so as to remove low harmonic numbers. In both cases, listeners fail to hear the two underlying pitches but instead report a single pitch corresponding roughly to that of the higher-rate pulse train (Carlyon, 1996; Carlyon *et al.*, 2002). Indeed, contrary to the effects seen with resolved harmonics, listeners do not consistently report pairs of such pulse trains that have widely different rates as sounding less fused than pairs whose rates are more similar (Carlyon, 1996). This suggests that, when two voices interact within a single channel, there should be little potential for implant users to exploit  $F_0$  differences for concurrent sound segregation.

The situation is slightly more encouraging when pairs of unresolved complex tones are filtered into *separate* frequency regions (Carlyon, 1994). Such a situation might arise when two speakers utter voiced sounds with different  $F_0$ s, and which, fortuitously, contain formants that occupy distinct and well-separated frequency regions. Listeners can detect  $F_0$  differences between such complexes (Carlyon, 1994), and Darwin (1992) has shown that across-frequency differences in  $F_0$  can cause a formant to “pop out” from the remainder of a voiced syllable, even when the harmonics of that formant is unresolved. Hence, source segregation via  $\Delta F_0$  may be possible provided the unresolved harmonics excites different regions of the basilar membrane. However, it should be noted that in Darwin’s study the popping out of a formant consisting of unresolved harmonics required a much larger  $\Delta F_0$  than when the harmonics was resolved, and, importantly, there was no evidence that this popping out affected the phonetic identity of the remainder of the syllable.

In the present study we investigated the extent to which, in the absence of resolved harmonics, normally hearing listeners can use pitch cues to segregate concurrent speech sounds. Specifically, we investigated the use of these “purely temporal” pitch cues under conditions where two competing sources occupied either completely overlapping or nonoverlapping, interleaved frequency ranges. To do so, we modified the noise-vocoding simulation of cochlear implant speech processors by replacing the noise-band carriers with bands of unresolved harmonics. This resulted in a waveform in each frequency channel that resembled a modulated pulse train. Spectral overlap of the two input waveforms was then manipulated by either processing both sentences on all six channels of the simulation and mixing them together, or presenting one sentence on the odd-numbered channels alone and the other on the even-numbered channels. The temporal pattern of stimulation was manipulated by using carrier harmonic complexes with either the same or different  $F_0$ s for the two speech sounds. In the simulation, the pitch cues were provided by the carrier  $F_0$ s used for each source, which were constant during a given sentence. This differs from previous approaches in which the “true”  $F_0$  of a single voice was tracked on a moment-by-moment basis and reflected in the rate of trains of pulses or noise bursts (Blamey *et al.*, 1984; Faulkner *et al.*, 2000). By holding pulse rate constant

for a given source, we aimed to maximize the chances of observing any effects of a rate difference between sources on concurrent sound segregation (Carlyon *et al.*, 2000). This also allowed us greater control of the resolvability of the pulse train, and hence of the validity of the simulation as a model of electric hearing.

The present study had two aims. First, we wished to determine the extent to which the psychophysical findings previously obtained with rather simple stimuli generalized to the perceptual segregation of competing sentences. Specific predictions arising from these findings are described in Sec. IV A, which also describes the conditions of our main experiment in more detail. Second, we wished to investigate whether, and how, implant users might be able to segregate concurrent sounds. The processing scheme we used took as its input two sources that have already been separated, and so could not, by itself, be implemented in a real-world device. Section V C briefly discusses one way in which automatic source segregation might be achieved. More generally, by gaining control of the temporal representation of competing speech sounds in the auditory periphery, the present experiments probe the temporal-processing limitations of the auditory system when resolved harmonics are absent.

## II. DESCRIPTION AND VALIDATION OF SIGNAL-PROCESSING TECHNIQUE

### A. Overview

The aim of our new signal-processing technique was to introduce a pitch cue that was encoded by purely temporal means, analogous to that produced by the pulse rate in a cochlear implant (Carlyon *et al.*, 2002). In particular, we wished to exclude any “place of excitation” cues to pitch. To do so, we modified the popular “noise vocoder” simulation of cochlear implant speech processors (Shannon *et al.*, 1995) by replacing the noise carrier with a harmonic complex (Fig. 1). Although the raw waveform of harmonic complex tones can, depending on the phase spectrum, resemble a pulse train, filtering by the normal cochlea can result in the place of excitation cues that we wished to avoid. To overcome this, a number of further modifications had to be made.

Place of excitation cues can be minimized by passing a complex with a low fundamental frequency ( $F_0$ ) through bandpass filters having relatively high center frequencies, at which auditory filters are broadest (Shackleton and Carlyon, 1994). Psychophysical experiments using this approach have yielded results that are similar to those obtained with single-channel cochlear implant simulations (Carlyon, 1996; McKay and Carlyon, 1999; Carlyon and Deeks, 2002; Carlyon *et al.*, 2002). We therefore used  $F_0$ s of (in different conditions) 40 and 70 Hz, and did not use any analysis channels having center frequencies (CFs) below 1089 Hz (Table I). Each of these manipulations had a potentially undesirable consequence.

First, the use of low  $F_0$ s meant that the “pulse rate” in the carrier signal might be too low to adequately sample the signal envelope. To alleviate this, harmonics were summed in alternating phase (Patterson, 1976; Shackleton and Carlyon, 1994), leading to pulse rates of 80 and 140 pps (double

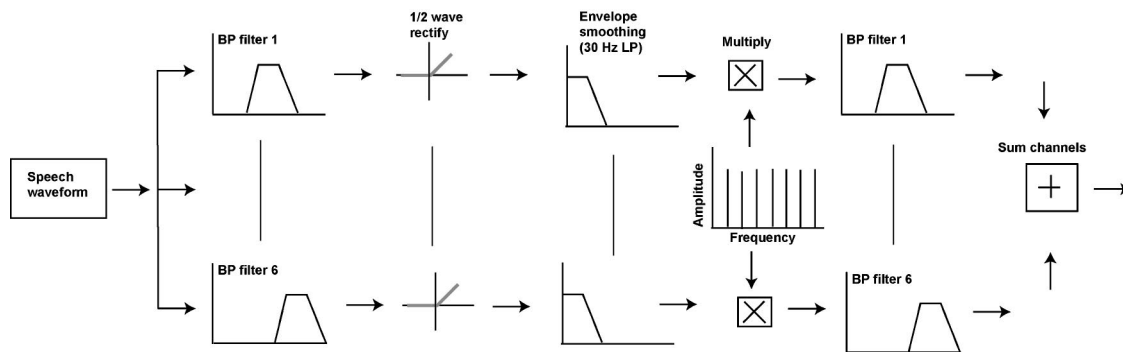


FIG. 1. Processing of ASL stimuli into experimental stimuli. The sentence waveform is analyzed using three or six bandpass filters. The output of each of these is half-wave rectified and smoothed using a 30-Hz low-pass filter. The resulting waveform is then multiplied with the carrier (a harmonic source with rate=80 or 140 pps) before passing through the bandpass synthesis filters. These outputs are then summed.

the respective  $F_0$ s). The signal envelope in each band was low-pass filtered at 30 Hz (4th-order Butterworths, giving 24-dB/octave attenuation rate), to reduce any aliasing effects in sampling the envelope even at the lower pulse rate. This also had the effect of eliminating any envelope fluctuations at rates equal to the  $F_0$  of the input voice, so the dominant “low” pitch of the complex was determined by the carrier pulse rate. The bandpass filters used for each channel were sixth-order Butterworths, with attenuation rates outside the passband of 18 dB/octave.

Second, eliminating analysis filters below 1089 Hz meant that we either had to discard speech information below that frequency, or spectrally shift the speech by introducing a mismatch between the analysis and synthesis bandpass filter in each channel. We chose the former option, as informal listening revealed that it produced a much smaller decrement in performance (cf. Fu and Shannon, 1999; Baskent and Shannon, 2003).<sup>1</sup> The root-mean-square levels in each band of the processed stimuli were set to the same level, thereby “whitening” the spectrum, and the overall level of the processed speech was 57.3 dB SPL. Finally, we added a continuous low-pass noise in order to mask distortion products having a frequency equal to the pulse rate (Pressnitzer and Patterson, 2001). It was generated by passing a white noise through a low-pass filter (Kemo VBF25.03; attenuation 48 dB/octave outside passband) having a corner frequency of 400 Hz and a spectrum level, within its passband, of 27.9 dB SPL.

To illustrate the temporal pitch code introduced by the scheme, a white noise was processed in the same way as the speech stimuli in the main experiment, with a pulse rate of

80 pps. The output was then passed through the peripheral stages of the Auditory Image Model (Patterson *et al.*, 1995). The resulting basilar-membrane motion (BMM), which demonstrates the effects of gammatone filtering, is shown in Fig. 2(a). It can be seen that all channels show a periodicity at a rate of 80 pps. Figure 2(b) shows the BMM in response to a condition, described in Sec. III B, in which the even-numbered channels of the processor output were deleted.

## B. Pitch preference judgments for sine- and alternating-phase single-channel stimuli

To check that the pitch conveyed by the pulse trains produced by the signal processing was purely temporal in origin, a preliminary control experiment was performed. All stimuli in this preliminary experiment were processed in the same way as the speech stimuli in the main experiment, except only the lowest-frequency channel, where resolved harmonics are most likely to occur, was used (see Table I). In addition, the harmonics used for the carrier could be summed in either alternating (ALT) or sine (SIN) phase. The input was always 0.5 s of white noise, with 20-ms raised-cosine onset and offset ramps. Each trial involved the presentation of three stimuli, the first of which was always an ALT-phase stimulus with  $F_0$  of either 40 or 70 Hz. The other two stimuli were always in SIN phase; one had an  $F_0$  equal to that of the first sound, and one had an  $F_0$  an octave higher. Shackleton and Carlyon (1994) have shown that, if the harmonics are unresolved, subjects should select this latter stimulus as having a pitch more like the first; in contrast, if the harmonics are resolved, the SIN-phase stimulus having an  $F_0$  equal to that of the first sound should be judged as more similar.

Table II shows the percentage of trials on which each subject reported that the ALT-phase complex had a pitch more like that of the SIN-phase complex having an  $F_0$  one octave higher. The results clearly show that the pitch arising from the lowest channel of the processed stimuli is one octave above the  $F_0$  of the complex, corresponding to a repetition rate of 80 or 140 pulses per second. In the remainder of this article we describe the carrier in terms of its pulse rate, rather than its  $F_0$ .

TABLE I. Frequency properties for each channel used in the simulations (kHz). The last column shows the number of components in the passband of the channel when  $F_0 = 70$  Hz.

Band	High-pass	Center	Low-pass	Bandwidth	Comps in BW
1	0.937	1.089	1.261	0.324	4
2	1.261	1.457	1.680	0.419	5
3	1.677	1.933	2.221	0.544	7
4	2.221	2.549	2.922	0.701	10
5	2.922	3.346	3.828	0.906	12
6	3.828	4.376	5.000	1.172	16

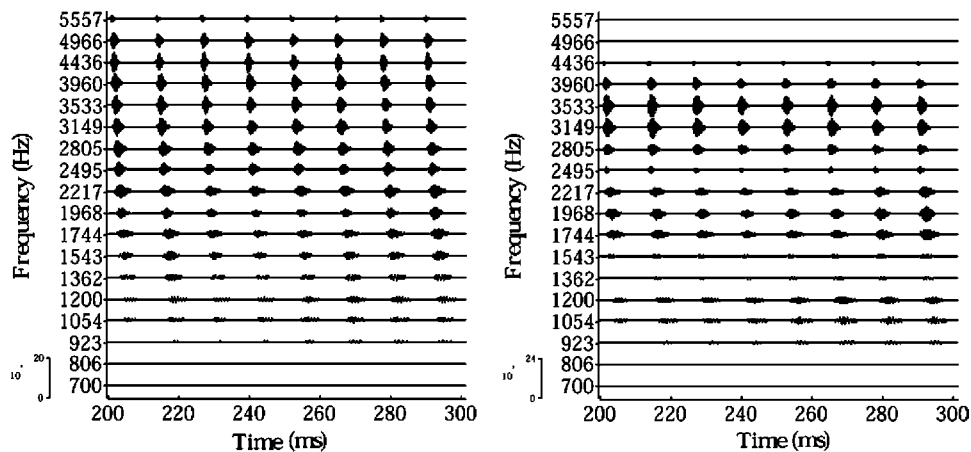


FIG. 2. Simulation of basilar-membrane motion (Patterson *et al.*, 1995) in response to white noise. (a) with all channels active. (b) with only odd-numbered channels active.

### III. EXPERIMENT 1. PERFORMANCE WITH A SINGLE SENTENCE AS A FUNCTION OF STIMULATION RATE AND NUMBER OF CHANNELS

#### A. Rationale

The purpose of experiment 1 was to investigate the effects of two manipulations—pulse rate and number of channels—on the perception of speech processed using the new scheme. These data were then used when interpreting the effects of experiment 2, which involved mixing target and masker sentences with the same or different pulse rates, and under conditions where the number of target channels differed.

#### B. Method

##### 1. Subjects

Sixteen normally hearing listeners (11 females and 5 males) took part. Thresholds in quiet for pure tones at 0.5, 1, 2, and 4 kHz were less than 20 dB HL in both ears for all subjects, except for two subjects with thresholds of 27.1 and 28.1 dB HL at 2 kHz in one ear. All subjects had English as their first language. Ages ranged from 18 to 36 years. All had some prior experience with listening to distorted speech sounds, but not with the type involved in this experiment.

##### 2. Stimuli and procedure

The speech material was taken from the MRC Institute of Hearing Research Adaptive Sentence Lists (ASL) (MacLeod and Summerfield, 1990). These sentences were based on the BKB sentences (Bench and Bamford, 1979) and produced by a male speaker of southern British English. Each sentence was scored for three keywords, using the “loose” scoring technique (Bench and Bamford, 1979). Each

TABLE II. Percentage of trials in which the ALT-phase  $F_0$  complex was judged to have a pitch more like the SIN-phase  $2F_0$  complex than that of the SIN-phase  $F_0$  complex.

Subject	Reference $F_0$ (Hz)	
	40	70
1	96.67	98.33
2	96.67	100.00
3	100.00	100.00

subject was presented with one sentence list per condition. As there were 15 sentences in each list, and 3 keywords per sentence, this yielded a score out of 45 for each condition.

The task for all subjects was to listen to each processed sentence and report (via keyboard entry) as many words as they could. Subjects clicked a button to move on to the next trial. Conditions were tested in blocks of 15 sentences, with short breaks taken between blocks. Testing took place in a double-walled sound-attenuating booth containing headphones, a mouse, and keyboard, and within sight of a large monitor.

The experiment measured speech reception performance with respect to combinations of two factors: speech processed into three or six channels, and with stimulation rates of 80 or 140 pps. Subjects were tested in a repeated-measures design across the four conditions, and were randomly assigned to one of four groups. The order of testing across these groups was counterbalanced. For three-channel conditions, half the subjects in each group were tested on odd-channel simulations, and half on even-channel simulations. The sentence lists used were also counterbalanced across conditions and groups. Hence, each group experienced a different sentence list for each condition, but, averaged across groups, each sentence was used an equal number of times for every condition. This was done to ensure that any differences in performance across condition were due to the processing rather than to any coincidental differences in difficulty between sentence lists. (Although the ASL lists are equated for difficulty across sentences, this equating was not performed with the type of processing used here.) The order of sentence presentation was randomized in each block for every subject and condition.

#### C. Results

Figure 3 shows the group mean percent correct for each condition. For the three-channel conditions, data from the odd- and even-channel subgroups were averaged, as they produced very similar results (odd vs even=35% vs 32% at 80 pps and 48% vs 43% at 140 pps).

Overall, performance was better for the six-channel conditions than for the three-channel conditions, and better at a rate of 140 pps than of 80 pps. These trends were confirmed by a repeated-measures ANOVA, having within-subject fac-

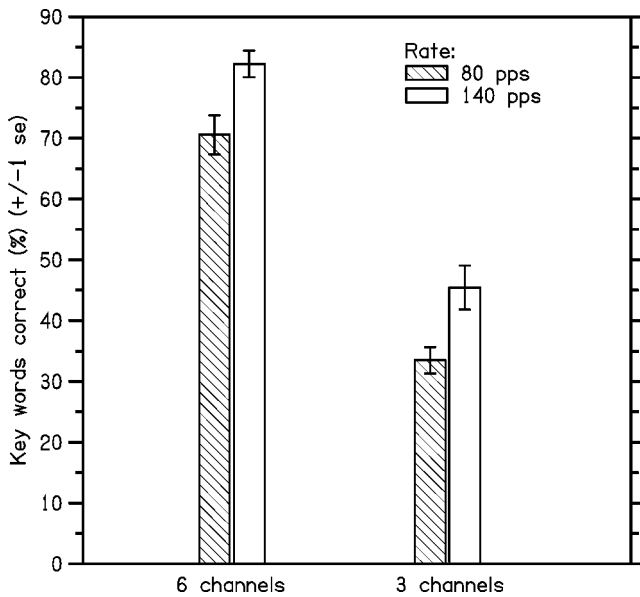


FIG. 3. Group mean scores (%) for single sentence conditions ( $\pm 1$  standard error).

tors of channel (three or six channels) and rate (80 or 140 pps), and a between-subject factor of group (odd or even channel subgroups for the three-channel conditions). Main effects of channel ( $F_{(1,14)}=246.2$ ,  $p<0.001$ ) and rate ( $F_{(1,14)}=25.0$ ,  $p<0.001$ ) were significant. There were no significant interactions and no effect of group.

#### D. Discussion

The experiment showed that higher carrier rates led to significantly higher speech reception scores. There have been no previous studies investigating the effect of pulse rate with normally hearing listeners and acoustic simulations, but a number of researchers have investigated the effect in co-

chlear implant listeners. Although the majority of these studied used pulse rates that exceeded the highest value of 140 pps employed here, Fu and Shannon (2000) investigated the range between 50 and 500 pps in four users of the Nucleus 22 implant. They implemented a four-channel CIS strategy, and, consistent with our results, found that phoneme recognition improved as stimulation rate was increased from 50 to 150 pps. Performance did not improve with further increases in rate, although Loizou *et al.* (2000) found that consonant and word recognition improved with increases in rate from 400 to 2100 pps.

Experiment 1 also showed that performance deteriorated as the number of channels was reduced from six to three. This was implemented by dropping alternate channels [Fig. 4(a)], unlike previous studies, in which a reduction in channel number was produced by broadening the analysis and carrier filters [Fig. 4(b)]. This effectively leads to “holes” in the speech spectrum, and the drop in performance can be attributed to the resulting loss of information, rather than to a loss of spectral resolution. In addition, it may be that some central limitation prevented subjects from effectively combining information across spectral bands separated by gaps. This possibility is suggested by the finding that introducing noise into (albeit much wider) spectral gaps can improve sentence recognition scores (Warren *et al.*, 1997).

## IV. EXPERIMENT 2: SENTENCE SEGREGATION USING PULSE-RATE AND CHANNEL DIFFERENCES

### A. Rationale and overview

Experiment 2 studied the ability of listeners to use pulse-rate and/or channel differences in segregating concurrent sentences. It used “target” sentences similar to those of experiment 1, and added them to a masker which started 1.2 s before the target. The target and masker were processed on

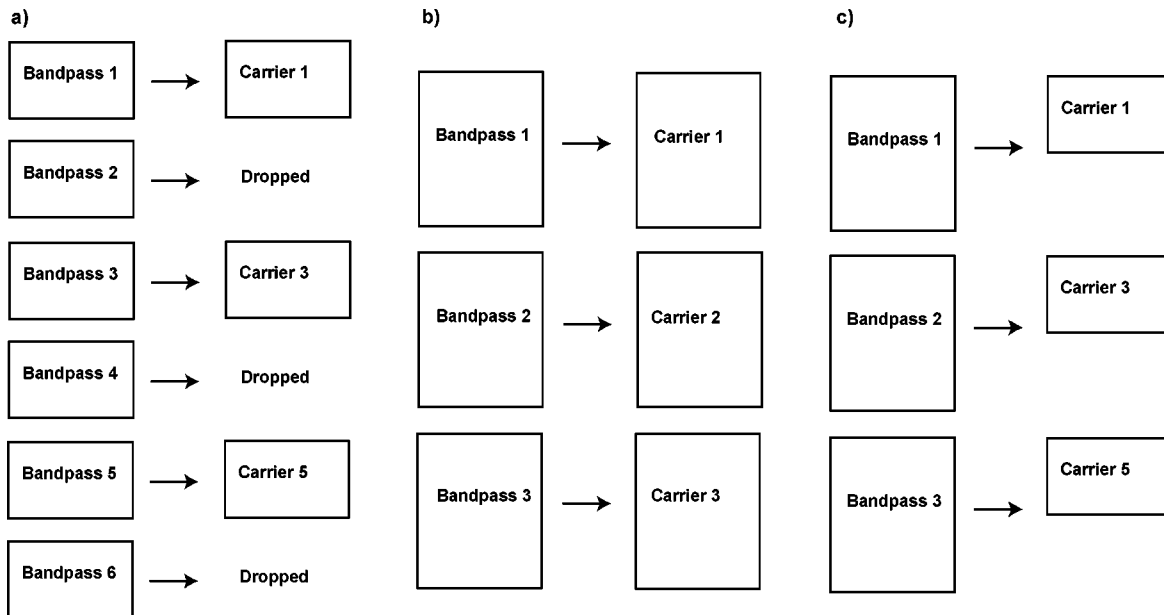


FIG. 4. Relation between analysis filters and carriers in three processing schemes. (a) Method used in this experiment, whereby alternate channels are dropped. (b) Conventional method, whereby fewer channels are accompanied by wider analysis filters. (c) Alternative method, whereby the entire frequency range is analyzed but still only narrow carrier bands are used (resulting in spectral mismatch).

all six channels and then mixed, or presented on three interleaved channels each. Orthogonal to this manipulation, the target and masker could be processed on either the same or different pulse rates.

As noted in the Introduction, previous psychophysical research suggests two ways in which rate differences could affect performance. First, when two equal-level pulse trains are mixed into the same channel, listeners hear a pitch equal to that of the higher-rate train (Carlyon *et al.*, 2002; van Wieringen *et al.*, 2003). Combined with the fact that the target sentence was added to the masker after some delay, this leads to an interesting prediction. When the masker is processed at a rate of 80 pps, introducing a 140-pps target should cause a change in pitch, perhaps aiding segregation of the two sources. In contrast, when the masker is processed at 140 pps, introducing an 80-pps target may leave the pitch unchanged, at least when the masker and target have the same level. This suggests that, when the target and masker levels are equal, introducing a rate difference between them should only help when the target is processed at the higher rate. In fact, as discussed below, the masker and target levels varied over time, and the target was on average more intense than the masker. Hence, as discussed further in Sec. V B, we might expect some advantage of a rate difference when the target is processed at the lower rate, but we would still expect the effects of a rate difference to be smaller than when the target is on the higher rate.

A second possible effect of a pulse rate difference occurs when the target and masker are processed on different channels. As discussed in the Introduction, it is possible that listeners could exploit a rate difference by comparing the temporal pattern of stimulation across different regions of the basilar membrane (Darwin, 1992; Carlyon, 1994). We would expect segregation based on such a cue to be symmetric, in that it should not depend on whether the target is processed at the higher or at the lower rate.

## B. Method

### 1. Stimuli

The sentence material was again taken from the ASLs. Target sentences were taken from different lists than the previous experiment, and were mixed with maskers. The maskers were constructed from three concatenated ASL sentences, which were time reversed and truncated to a duration of 3.5 s, with 20-ms Hanning onset and offset ramps. The type of masker was chosen so as to have many of the spectral and temporal characteristics of interfering speech, while itself being unintelligible. This latter characteristic avoided problems associated with scoring responses that corresponded either to whole words in a competing sentence, or to composites of words from a target and competing sentence. None of the sentences used to construct the maskers had been used in the first part of the experiment, and each target sentence was combined with a masker constructed from a unique set of three ASL sentences.

Both the target and masker were processed in the same way as described in the previous experiment. They were each separately processed into three or six channels with the de-

TABLE III. Parameters used in conditions 1 to 4 of experiment 2.

Condition	Target sentence		Masker sentence	
	Rate	Chans	Rate	Chans
1a	80	All	80	All
1b	140	All	140	All
2a	80	All	140	All
2b	140	All	80	All
3a	80	Odd	80	Even
3b	80	Even	80	Odd
3c	140	Odd	140	Even
3d	140	Even	140	Odd
4a	80	Odd	140	Even
4b	80	Even	140	Odd
4c	140	Odd	80	Even
4d	140	Even	80	Odd

sired carrier rate (depending on the condition). The masker was then attenuated by 9 dB relative to the target sentence, and the target was added to the masker with an offset of 1.2 s. Because this value is an integer multiple of the periods of the 80-pps and 140-pps pulse trains, the pulses from each source were simultaneous when the target and masker were processed on the same rate. This gave rise to a single pulse train in each frequency region, that was modulated by an envelope derived from the sum of the filtered masker and target waveforms. All stimuli were generated at a sample rate of 22 050 Hz with 16-bit resolution.

Stimulus presentation was through the same headphones as for experiment 1. Subjects were again seated in the double-walled sound-attenuating chamber containing the headphones, keyboard, and mouse, within sight of the computer monitor. They were instructed to report back as much of the target sentence as they could. They were informed that the target sentence would start later than the interfering sound, and that it might be accompanied by an increase in loudness.

### 2. Subjects, design, and procedure

The experiment investigated factors of number of channels (all six, and odd- or even-channels only) and rate (80 and 140 pps) with respect to target and masker. The same subjects who participated in experiment 1 took part.

Table III shows the structure of conditions. In condition 1, both target and masker were processed into six channels, and both had the same carrier rate (either 80 or 140 pps). In condition 2, both target and masker were processed into six channels, but this time had different carrier rates (target=80 pps; masker=140 pps, or vice versa). Condition 3 involved simulations with targets and maskers processed on odd or even channels only (target=odd, masker=even, or vice versa). Both target and masker had the same carrier rate, which was either 80 or 140 pps. Condition 4 involved simulations with targets and maskers having different channels and carrier rates.

Subjects were randomly assigned to one of four groups. For each group, a different set of sentences was used for every condition. However, by counterbalancing the alloca-

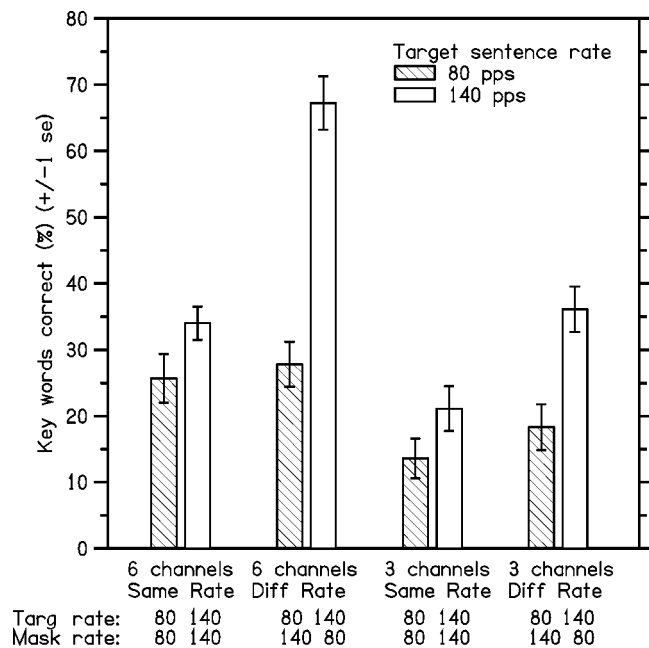


FIG. 5. Group mean scores (%) for target and interferer sentence conditions ( $\pm 1$  standard error).

tion of sentence lists to condition across groups, each sentence was used an equal number of times (although for different subjects) for each condition. The order in which the different conditions were run was also fully counterbalanced across groups. Subjects were tested in a total of eight conditions. All 16 subjects performed conditions 1a, 1b, 2a, and 2b. For conditions 3 and 4, subjects were further divided into two subgroups, depending on whether odd- or even-numbered channels were used for the target. Hence eight subjects did conditions 3a, 3c, 4a, 4c, and eight did 3b, 3d, 4b, and 4d.

### C. Results

As with experiment 1, differences between performance in odd- and even-channel subgroups in the three-channel conditions were small, and data were averaged across the odd- and even-channel sub-groups. Figure 5 shows the group mean and standard error for each condition. Overall, performance is worse than that found for single-sentence tasks (Fig. 3). The remaining trends will be discussed with reference to the results of a repeated measures ANOVA applied to the data. The ANOVA had within-subject factors of channels (three vs six), target rate (80 vs 140 pps), and rate difference (target rate same or different to masker rate). It also used a between-subjects factor of group (odd or even channels for three-channel conditions), which was found not to be significant.

As expected from the results of experiment 1, performance was better at a target rate of 140 pps than of 80 pps (open vs filled bars:  $F_{(1,14)} = 97.0$ ,  $p < 0.001$ ). It was also better for six than for three channels ( $F_{(1,14)} = 60.3$ ,  $p < 0.001$ ), indicating that any effect of spectral separation between the target and masker channels was swamped by the reduction in conveyed information produced by deleting alternate channels. This latter reduction was substantial, pro-

ducing an average decrease in performance from 37% to 23% for masked speech (experiment 2), compared with 76% to 39% correct for speech in quiet (experiment 1).

Performance was also better overall when the target and masker were processed on different rates ( $F_{(1,14)} = 39.3$ ,  $p < 0.001$ ). This difference occurred only when the target was processed on the higher rate (open bars), leading to an interaction between target rate and rate difference ( $F_{(1,14)} = 26.74$ ,  $p < 0.001$ ).<sup>2</sup> As was discussed in Sec. IV A, this is consistent with the psychophysical evidence that, when two pulse trains interact in the same channel, the pitch is determined by the higher of the two pulse rates. Two further points are worth making about this interaction. First, it was not the case that introducing a rate difference *decreased* performance when the target was processed on 80 pps, as would have occurred if the 140-pps stimulus simply produced more excitation in the auditory periphery. Hence, by introducing a rate difference, it is possible to improve performance when the target is processed on the higher rate, without any cost when processed on the lower rate—at least for conditions like those used here, where the target was more intense than the masker. Second, it occurred not only when the target and masker were presented on the same six channels, but also when they were processed on three interleaved channels each. Specifically, a rate difference failed to improve performance in the three-channel condition when the target was processed on the lower rate. Hence, there is no evidence for the symmetric improvement caused by introducing a rate difference that one might expect if an across-frequency rate comparison aided the perceptual segregation of concurrent speech sounds (Sec. IV A).

The effect of a rate difference when the target and masker were presented on separate channels was qualitatively similar to that observed when they shared all six channels. This may have been due to spread of excitation on the basilar membrane, causing the responses to target and masker to overlap partially. This is illustrated in Fig. 2(b), which shows the output of a bank of auditory filters in response to a white noise processed only on the three odd-numbered channels. It can be seen that although there was no energy in the even-numbered channels of the stimulus, there is some activity in the outputs of auditory filters having CFs within the passbands of those channels. For example, auditory filters centered on 2805 and 3060 Hz both fall within the passband of even-numbered channels (Table I), yet show a nonzero output. This issue is discussed further in Sec. V C.

Finally, we should note that the effect of a rate difference was actually *smaller* in the three- than in the six-channel conditions—as revealed by a significant interaction between the effects of a rate difference and of the number of channels ( $F_{(1,14)} = 25.84$ ,  $p < 0.001$ ). We can think of two reasons for this interaction. First, it may be an artifact of the finding that performance was lower overall in the three-channel conditions. It could be that the improvement from 17% to 27% (averaged across rate) in the three-channel condition actually reflected a similar change in underlying sensitivity as the numerically larger increase (from 30% to 48%) in the six-channel condition. This would occur, for example, if low levels of performance were partially affected by a

floor effect. If this were true, we might expect across-subject standard deviations to be smaller over this range. To test this, we calculated the regression between standard deviation and percent correct for all subjects and all conditions of experiments 1 and 2. The resulting regression had a slope of only  $-0.004$  ( $r^2=0.02$ ), which was not significantly different from zero—arguing against a floor effect. An alternative explanation is that, when the competing stimuli were filtered into interleaved channels, there was less interaction between them. Because the dominance of the pitch of higher-rate pulse trains has only been observed when the components fall within a single channel (Carlyon *et al.*, 2002; van Wieringen *et al.*, 2003), this effect would be expected to depend on the extent of such within-channel interactions. Where such interactions are reduced, whatever segregation does occur might be expected to depend on spectral separation between the masker and target, rather than on differences in pulse rate.

## V. DISCUSSION

### A. Overview

This article presents results from a new signal-processing strategy that reproduces several features of the peripheral pattern of stimulation produced by cochlear implants. In common with previous simulations (Shannon *et al.*, 1995) it extracts the envelope in a number of fairly broad frequency bands, and uses each envelope to modulate a carrier signal having an appropriate frequency content. The novel aspect lies in being able to modify the temporal aspects of the carrier, and, as our preliminary experiment showed, to do so without introducing resolved frequency components. This allowed us to study the effects of varying the “pulse rate” in each channel, and of introducing a difference between the pulse rates applied to competing sources. It resulted in a number of new findings, some of which have implications for concurrent sound segregation in general, and, in particular, for how this may be achieved by cochlear implant users. Furthermore, some aspects of the data on the effects of pulse rate differences between the target and masker were interpreted in terms of recent evidence on the basic psychophysics of “purely temporal” pitch perception.

### B. Effects of pitch differences and neural refractoriness

#### 1. Comparison to previous data on temporal pitch perception

One of the most important findings from the present study is that processing a masker on a different rate from the target aids performance only when the target rate is higher than the masker rate. As discussed in Sec. IV A, this is roughly consistent with psychophysical evidence that, when two pulse trains of rates  $R_1$  and  $R_2$  are mixed on the same channel in acoustic or electric hearing, subjects hear a single pitch roughly equal to  $R_2$  (Carlyon, 1996; Carlyon *et al.*, 2002; van Wieringen *et al.*, 2003). This should result in a large increase in pitch when a 140-pps target is added to an 80-pps masker (condition “T140/M80:” new pitch about 140 pps), but a much smaller (or no) change when an 80-pps

target is added to a 140-pps masker (T80/M140: pitch stays close to 140 pps). One caveat is worth mentioning when relating the two sets of findings: The pairs of pulse trains that were mixed together in the psychophysical experiments had equal levels, whereas here the level of the target, averaged over its total duration, was 9 dB more intense than that of the masker. Presumably, there will be some SNR at which the target will dominate the pitch even when it is processed at a lower rate than the masker. However, it is worth noting that the levels of both target and masker varied throughout the utterances, and so there will have been instances where the SNR was lower than 9 dB (and some where it was higher). Hence, the psychophysical data would predict a trend in the direction observed here, although it is perhaps a little surprising that a rate difference had no beneficial effect at all when the target was processed on the lower rate.

## C. Implications for concurrent sound segregation

### 1. Overview

As pointed out in the Introduction, cochlear implant users currently experience particular difficulty in understanding speech when competing sounds are present. Part of this problem may well be due to the incomplete survival of peripheral processes. However, because similar limitations are observed with acoustic simulations of cochlear implant speech-processing algorithms, it seems likely that these algorithms do not faithfully transmit cues that are important for concurrent sound segregation. As discussed in the Introduction, pitch cues are likely to fall into this category.

There are three broad classes of potential solution to this problem. First, it is possible that, without explicitly segregating the sources *a priori*, monaural cues to concurrent sound segregation could be introduced in new speech-processing strategies, perhaps combined with new electrode designs that restrict the spread of excitation along the neural array. For example, one could attempt to reintroduce the combined place-of-excitation and timing coding of resolved harmonics. However, as pointed out by Moore and Carlyon (2004), this seems unlikely to be achieved in the foreseeable future. A more promising and immediate solution is provided by bilateral implants. Although this approach does not currently allow significant use of interaural timing cues, substantial advantages can be produced by the head shadow effect (e.g., van Hoesel and Tyler, 2003). One drawback, however, is that bilateral implantation is unlikely to be economically justifiable in the medium term, at least in countries whose health-care system is primarily publicly funded (Summerfield *et al.*, 2002). As van Hoesel and Tyler (2003) have pointed out, the head-shadow effect raises the (cheaper) possibility of using two microphones to route two separate sources to a single implant. They also pointed out that such a strategy, unlike bilateral implants, would not allow the listener to select each source at will. Indeed, this is a problem for any scheme which performs automatic source segregation: for the listener to take advantage of it, the two sources must be encoded in a way which then allows them to be perceptually segregated, and for the listener to switch between sources as required. The present experiments implemented an acoustic analog of

one such encoding scheme, in which different sources are applied to different subsets of electrodes, and with different carrier rates applied to electrodes encoding different sources.

## 2. Across-frequency rate differences

We believe that the results of the present study should lead one to be cautiously pessimistic about the use of across-channel differences in pulse rate to allow perceptual segregation of different sound sources. The only significant advantage gained by presenting the target and masker on different rates occurred when the target was on the higher of the two possible rates. Performance was substantially better when the target was processed on 140 pps and the masker on 80 pps (T140/M80) than in condition T140/M140 (both processed on 140 pps). However, in a real-life device, where the target and masker are not specified beforehand, the choice is likely to be between processing the two sources on different rates or processing them both on the higher of the two rates. Our results indicate that the former option could lead to worse performance than the latter, as listeners were worse in condition T80/M140 than in condition T140/M140 (compare the third and second bars from the left in Fig. 5).

Experiment 2 showed only a small effect of rate differences when the target and masker were processed on separate channels. Furthermore, this advantage showed the same asymmetry as occurred when they were both processed on all six channels, consistent with it being due to spread of excitation between adjacent channels. The center frequencies of adjacent channels shown in Table I were separated by an average of about 23%. Hence, an auditory filter centered on one frequency band would respond to components in the middle of a neighboring band with an attenuation of approximately 30 dB (Patterson *et al.*, 1982). However, there would also have been auditory filters centered on the boundary between two channels, and, when the masker and target were interleaved, these would have responded equally to the two sources.

Although we cannot rule out the possibility that a wider channel separation would have allowed subjects to use across-channel rate differences more effectively, we think it unlikely that such a process could aid concurrent sound segregation in existing cochlear implants. The rate differences used here were large and constant, and yet the pattern of results was consistent with a purely within-channel effect. Unless an implant were able to produce substantially more effective between-channel attenuation than the 30 dB in the present study, it seems unlikely that, even if across-channel rate differences were usable, this could effectively be realized. This stands in marked contrast to the situation in normal acoustic hearing, where resolved harmonics are present, in which case such across-channel processes can have a marked effect (Broadbent and Ladefoged, 1957; Darwin, 1992).

## 3. Assigning sources to different channels

An additional finding of experiment 2 was that performance was consistently worse when the target and masker were each presented to three (separate) channels than when they were mixed into the same six channels. As noted in Sec.

IV C, this indicates that any advantage of spectrally separating the two sources was swamped by the deleterious effects of reducing the number of channels per source. One might be tempted to conclude that any attempt to separate two sources on a channel-by-channel basis is doomed to failure, because it would reduce the number of channels for each source. However, there are two reasons why this is not necessarily so. First, the effects of halving the number of channels should decrease, the more channels there are to start with (Friesen *et al.*, 2001). Second, the effects of halving the number of channels may be reduced by adopting a slightly different method of splitting the two sources than the one used here [Fig. 4(a)]. Note that the “standard” method, of doubling the bandwidths of both the analysis and carrier channels [Fig. 4(b): Shannon *et al.*, 1995], would not be appropriate, as alternate channels would be needed for separate sources. However, one could double the bandwidth of the analysis channels, while using the same (narrow) carrier channels as in the present scheme—as shown in Fig. 4(c). We intend to explore this issue in a future study.

## 4. Conclusions

In summary, the absence of an advantage produced by a difference in pulse rate when the target is presented on the lower of two rates suggests that rate differences are unlikely to provide a consistent cue for segregation produced by a real-world device. However, although the present study showed no advantage of spectrally separating the two sources, it is possible that such an advantage could be observed when more channels are available. For implant users to take full advantage of such a strategy, it is likely that they would have to exploit cues to concurrent sound segregation other than differences in pulse rate used here. Such cues might include the natural across-frequency covariations in amplitude present in each source.

## VI. SUMMARY

- (i) We have described a new simulation of cochlear implant hearing in which the envelope in each frequency band modulates a bandpass filtered pulse train. Because the frequency components of the carrier are unresolved in the auditory periphery, this allows one to simulate the effects of varying pulse rate in electric hearing. Pitch can be manipulated independently of the  $F0$  of the input.
- (ii) Performance is better when the target is processed at a rate of 140 pps than at 80 pps, and when processed on six than on three channels. Both of these findings occurred for speech in quiet and in the presence of a masker.
- (iii) Presenting the target on only the odd-numbered channels and the masker on even-numbered channels (or vice versa) produced worse performance than processing them both on all six channels. Hence, for the particular stimuli used here, any benefits of spectral separation were outweighed by the reduction in number of channels conveying the target.

- (iv) There was no evidence that listeners can exploit across-channel differences in pulse rate to identify a target sentence in the presence of a masker.
- (v) When the target and masker are processed on the same channels, and the target level is higher, performance can be improved by processing the masker on a lower rate than the 140-pps target (compared to when they are processed on the same rate). However, processing the masker on a higher rate than the 80-pps target neither helps or hinders, again compared to the case where both are processed on the same rate.

## ACKNOWLEDGMENTS

The signal-processing scheme used here was based on a MATLAB implementation of the noise-vocoder algorithm, based on code written by Philipos Loizou and subsequently modified by Stuart Rosen. We thank Johannes Lyzenga for further modification of the code, and for advice on its implementation. We also thank Johannes Lyzenga, Christopher Long, and Fan-Gang Zeng for helpful comments on an earlier draft of this manuscript.

<sup>1</sup>Recent evidence has shown that speech perception in noise can be improved by presenting *unprocessed* low-frequency sound to implant users having low-frequency hearing in either the implanted or contralateral ear (Kong *et al.*, 2003; Turner *et al.*, 2003). However, performance is poor when the low-frequency information is presented via an implant or an acoustic simulation thereof (Nelson and Jin, 2002; Kong *et al.*, 2003), presumably due to the lack of resolved harmonics (see the Introduction). It therefore seems unlikely that low-frequency *information* plays an especially important role in implants or in experiments such as ours, in which resolved harmonics are absent.

<sup>2</sup>A two-way ANOVA performed only on the data with the 80-pps target revealed an effect of number of channels, but no effect of having the target on the same vs a different rate ( $F_{(1,15)} = 1.39$ ,  $p = 0.26$ ). When a similar analysis was performed on the data with a 140-pps target, the effect of masker rate was highly significant ( $F_{(1,15)} = 71.56$ ,  $p < 0.001$ ).

Assmann, P., and Summerfield, Q. (1990). "Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies," *J. Acoust. Soc. Am.* **88**, 680–697.

Baskent, D., and Shannon, R. V. (2003). "Speech recognition under conditions of frequency–place compression and expansion," *J. Acoust. Soc. Am.* **113**, 2064–2076.

Bench, J., and Bamford, J. (1979). *Speech-hearing Tests and the Spoken Language of Hearing Impaired Children* (Academic, London).

Blamey, P. J., Dowell, R. C., Tong, Y. C., Brown, A. M., Luscombe, S. M., and Clark, G. (1984). "Speech processing strategies using an acoustic model of a multiple-channel cochlear implant," *J. Acoust. Soc. Am.* **76**, 104–110.

Broadbent, D. E., and Ladefoged, P. (1957). "On the fusion of sounds reaching different sense organs," *J. Acoust. Soc. Am.* **29**, 708–710.

Brokx, J. P. L., and Nootboom, S. G. (1982). "Intonation and the perceptual separation of simultaneous voices," *J. Phonetics* **10**, 23–36.

Brown, G. J., and Cooke, M. (1994). "Computational auditory scene analysis," *Comput. Speech Lang.* **8**, 297–336.

Carlyon, R. P. (1994). "Detecting pitch–pulse asynchronies and differences in fundamental frequency," *J. Acoust. Soc. Am.* **95**, 968–979.

Carlyon, R. P. (1996). "Encoding the fundamental frequency of a complex tone in the presence of a spectrally overlapping masker," *J. Acoust. Soc. Am.* **99**, 517–524.

Carlyon, R. P., and Deeks, J. M. (2002). "Limitations on rate discrimination," *J. Acoust. Soc. Am.* **112**, 1009–1025.

Carlyon, R. P., Moore, B. C. J., and Micheyl, C. (2000). "The effect of modulation rate on the detection of frequency modulation and mistuning of complex tones," *J. Acoust. Soc. Am.* **108**, 304–315.

Carlyon, R. P., van Wieringen, A., Long, C. J., Deeks, J. M., and Wouters, J. (2002). "Temporal pitch mechanisms in acoustic and electric hearing," *J. Acoust. Soc. Am.* **112**, 621–633.

Culling, J. F., and Darwin, C. J. (1993). "Perceptual separation of simultaneous vowels—Within and across-formant grouping by  $F_0$ ," *J. Acoust. Soc. Am.* **93**, 3454–3467.

Darwin, C. J. (1992). "Listening to two things at once," in *Audition, Speech, and Language*, edited by B. Schouten (Mouton-De Gruyter, Berlin), pp. 133–148.

Faulkner, A., Rosen, S., and Smith, C. (2000). "Effects of the salience of pitch and periodicity information on the intelligibility of four-channel vocoded speech: Implications for cochlear implants," *J. Acoust. Soc. Am.* **108**, 1877–1887.

Friesen, L. M., Shannon, R. V., Baskent, D., and Wang, X. (2001). "Speech recognition in noise as a function of the number of spectral channels: Comparisons of acoustic hearing and cochlear implants," *J. Acoust. Soc. Am.* **110**, 1150–1163.

Fu, Q. J., and Shannon, R. V. (1999). "Recognition of spectrally degraded and frequency-shifted vowels in acoustic and electric hearing," *J. Acoust. Soc. Am.* **105**, 1889–1900.

Fu, Q. J., and Shannon, R. V. (2000). "Effect of stimulation rate on phoneme recognition by Nucleus-22 cochlear implant listeners," *J. Acoust. Soc. Am.* **107**, 589–597.

Geurts, L., and Wouters, J. (2001). "Coding of the fundamental frequency in continuous interleaved sampling processors for cochlear implants," *J. Acoust. Soc. Am.* **109**, 713–726.

Kong, Y.-Y., Stickney, G. S., and Zeng, F.-G. (2003). "Combined acoustic and electric hearing for speech in noise and melody recognition," 2003 Conference on Implantable Auditory Prostheses; Asilomar Conference Grounds, Pacific Grove, CA.

Loizou, P. C., Poroy, O., and Dorman, M. (2000). "The effect of parametric variations of cochlear implant processors on speech understanding," *J. Acoust. Soc. Am.* **108**, 790–802.

MacLeod, A., and Summerfield, Q. (1990). "A procedure for measuring auditory and audio-visual speech-reception thresholds for sentences in noise: Rationale, evaluation, and recommendations for use," *Br. J. Audiol.* **24**, 29–43.

McKay, C. M., and Carlyon, R. P. (1999). "Dual temporal pitch percepts from acoustic and electric amplitude-modulated pulse trains," *J. Acoust. Soc. Am.* **105**, 347–357.

Meddis, R., and Hewitt, M. (1992). "Modeling the identification of concurrent vowels with different fundamental frequencies," *J. Acoust. Soc. Am.* **91**, 233–245.

Moore, B. C. J., and Carlyon, R. P. (2004). "Perception of pitch by people with cochlear hearing loss and by cochlear implant users," in *Springer Handbook of Auditory Research: Pitch Perception*, edited by C. J. Plack and A. J. Oxenham (Springer, Berlin, in press).

Nelson, P. B., and Jin, S.-H. (2002). "Understanding speech in single-talker interference: Normal-hearing listeners and cochlear implant users," *J. Acoust. Soc. Am.* **111**, 2429.

Patterson, R. D. (1976). "Auditory filter shapes derived with noise stimuli," *J. Acoust. Soc. Am.* **59**, 640–654.

Patterson, R. D., Allerhand, M., and Giguère, C. (1995). "Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform," *J. Acoust. Soc. Am.* **98**, 1890–1894.

Patterson, R. D., Nimmo-Smith, I., Weber, D. L., and Milroy, R. (1982). "The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold," *J. Acoust. Soc. Am.* **72**, 1788–1803.

Pressnitzer, D., and Patterson, R. D. (2001). "Distortion products and the pitch of harmonic complex tones," in *Physiological and Psychophysical Bases of Auditory Function*, edited by D. J. Breebart, A. J. M. Houtsma, A. Kohlrausch, V. F. Prijs, and R. Schoonhoven (Shaker, Maastricht), pp. 97–104.

Qin, M. K., and Oxenham, A. J. (2003a). "The effects of simulated cochlear-implant processing on  $F_0$  discrimination," *J. Acoust. Soc. Am.* **113**, 2224.

Qin, M. K., and Oxenham, A. J. (2003b). "Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers," *J. Acoust. Soc. Am.* **114**, 446–454.

Scheffers, M. T. M. (1983). "Sifting vowels: Auditory pitch analysis and sound segregation," Doctoral dissertation, University of Groningen, Netherlands.

- Shackleton, T. M., and Carlyon, R. P. (1994). "The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination," *J. Acoust. Soc. Am.* **95**, 3529–3540.
- Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). "Speech recognition with primarily temporal cues," *Science* **270**, 303–304.
- Summerfield, A. Q., Marshall, D. H., Barton, G. R., and Bloor, K. E. (2002). "A cost-utility scenario analysis of bilateral cochlear implantation," *Arch. Otolaryngol. Head Neck Surg.* **128**, 1255–1262.
- Turner, C., Gantz, B., Lowder, M., and Gfeller, K. (2003). "Combining acoustic and electric hearing: Clinical studies," 2003 Conference on Implantable Auditory Prostheses; Asilomar Conference Grounds, Pacific Grove, CA.
- van Hoesel, R. J. M., and Tyler, R. S. (2003). "Speech perception, localization, and lateralization with bilateral cochlear implants," *J. Acoust. Soc. Am.* **113**, 1617–1630.
- van Wieringen, A., Carlyon, R. P., Long, C. J., and Wouters, J. "Pitch of amplitude-modulated irregular-rate stimuli in electric and acoustic hearing," *J. Acoust. Soc. Am.* **114**, 1516–1528.
- Warren, R. M., Riener Hainsworth, K., Brubaker, B. S., Bashford, J. A., and Healy, E. W. (1997). "Spectral restoration of speech: Intelligibility is increased by inserting noise in spectral gaps," *Percept. Psychophys.* **59**, 275–283.
- Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., Eddington, D. K., and Rabinowitz, W. M. (1991). "Better speech recognition with cochlear implants," *Nature (London)* **352**, 236–238.