Enhancement of temporal periodicity cues in cochlear implants: Effects on prosodic perception and vowel identification

Tim Green,[a] Andrew Faulkner, Stuart Rosen, and Olivier Macherey
Department of Phonetics and Linguistics, University College London, Wolfson House, 4 Stephenson Way, London, NW1 2HE, United Kingdom

(Received 27 July 2004; revised 22 March 2005; accepted 8 April 2005)

Standard continuous interleaved sampling processing, and a modified processing strategy designed to enhance temporal cues to voice pitch, were compared on tests of intonation perception, and vowel perception, both in implant users and in acoustic simulations. In standard processing, 400 Hz low-pass envelopes modulated either pulse trains (implant users) or noise carriers (simulations). In the modified strategy, slow-rate envelope modulations, which convey dynamic spectral variation crucial for speech understanding, were extracted by low-pass filtering (32 Hz). In addition, during voiced speech, higher-rate temporal modulation in each channel was provided by 100% amplitude-modulation by a sawtooth-like wave form whose periodicity followed the fundamental frequency ($F0$) of the input. Channel levels were determined by the product of the lower- and higher-rate modulation components. Both in acoustic simulations and in implant users, the ability to use intonation information to identify sentences as question or statement was significantly better with modified processing. However, while there was no difference in vowel recognition in the acoustic simulation, implant users performed worse with modified processing both in vowel recognition and in formant frequency discrimination. It appears that, while enhancing pitch perception, modified processing harmed the transmission of spectral information. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1925827]

PACS number(s): 43.66.Ts, 43.71.Bp, 43.66.Hg [KWG] Pages: 375–385

I. INTRODUCTION

As cochlear implant systems have becoming increasingly successful at providing basic speech recognition abilities, consideration has begun to be given to their performance in aspects of speech perception that have previously received little attention. One such aspect is intonation, as conveyed by voice pitch variation. Intonation is, of course, crucial to the perception of tonal languages in which pitch information conveys lexical meaning, but also has important functions in all other languages. It is a major component of prosody, and is widely held to play an important role in early language development (e.g., Jusczyk, 1997), which is of particular significance in light of the increasing number of very young children receiving implants.

In normal hearing, the principal cues to voice pitch derive from the ability of the auditory periphery to divide the speech input into a large number of frequency channels. Lower speech harmonics are resolved, and the neural coding of resolved harmonics provides both place and temporal cues related to $F0$. In contrast, in most commonly used cochlear implant speech processing strategies, spectro-temporal information is conveyed in the form of fixed-rate pulse trains that are amplitude modulated by low-pass envelopes extracted from a relatively small number of frequency bands (Seligman and McDermott, 1995; Wilson *et al.*, 1991; Vandali *et al.*, 2000). As a consequence, speech harmonics are not resolved, the spectral ("place") cues used in normal hearing are not available, and pitch perception is thought to depend upon the derivation of temporal cues from modulation components of the amplitude envelope that are related to voice $F0$. Such cues are, in principle, available, as long as the cutoff frequency of the envelope smoothing filter is high enough to pass $F0$, and the pulse rate is at least four to five times the modulation frequency, allowing accurate representation of the modulating envelope (McKay, McDermott, and Clark, 1994; Wilson, 1997). However, the ability to perceptually encode temporal amplitude modulation is limited, particularly for frequencies at the higher end of the human voice pitch range, both in normally hearing listeners presented with amplitude modulated noise (Burns and Viemeister, 1976; Burns and Viemeister, 1981; Formby, 1985; Grant, Summers, and Leek, 1998; Hanna, 1992; Pollack, 1969) and in implant users detecting amplitude modulations in pulse trains (Busby, Tong, and Clark, 1993; Cazals *et al.*, 1994; Donaldson and Viemeister, 2002; Shannon, 1992).

Consistent with these constraints on the availability of pitch cues, the limited amount of research examining voice pitch perception in continuous interleaved sampling (CIS) or similarly processed speech has shown levels of intonation perception that are severely limited compared to normal hearing. This is true both for implant users (Au, 2003; Barry *et al.*, 2002; Ciocca *et al.*, 2002; Green, Faulkner, and Rosen, 2004; Lee *et al.*, 2002; Peng *et al.*, 2004; Wei *et al.*, 2000; Wei *et al.*, 2004), and for normally hearing subjects listening to noise-excited vocoder acoustic simulations (Fu *et al.*, 1998; Green, Faulkner, and Rosen, 2002; Green, Faulkner, and Rosen, 2004; Xu, Tsai, and Pfingst, 2002).

[a]Electronic mail: tim@phon.ucl.ac.uk

In some early speech processing strategies, the rate of pulsatile stimulation was controlled by *F*0 (e.g., Fourcin *et al.*, 1979; Tong *et al.*, 1980). Implant users are generally able to discriminate differences in pulse rate reasonably well for rates up to a few hundred hertz (McDermott and McKay, 1997; Pijl and Schwarz, 1995; Townshend *et al.*, 1987; Wilson *et al.*, 1997; Zeng, 2002) and such an approach may well be optimal for conveying voice pitch information. However, such strategies are typically worse on standard measures of speech perception than CIS-like strategies, presumably because of poorer transmission of spectral information (e.g., McKay and McDermott, 1993; McKay *et al.*, 1992; Skinner *et al.*, 1999). Consequently, this kind of strategy has largely been discarded, although there has recently been some interest in the possibility that, for implant users who speak tonal languages, the advantages in tone recognition afforded may outweigh the potential disadvantages (Lan *et al.*, 2004). In general, however, for the foreseeable future, the majority of implant users are likely to use CIS-like strategies, particularly since most recent developments in the design of implant systems have been motivated by the possibility that very high pulse rates may lead to patterns of neural responses that more closely resemble those found in normal hearing (e.g., Rubenstein *et al.*, 1999).

One recent approach to enhancing pitch cues in CIS-processed speech involved modifying the properties of the analysis filters that divide the input into frequency bands, to try to restore a form of place cue to pitch (Geurts and Wouters, 2004). The filters typically used in CIS processing have flat passbands and are relatively broad, particularly at lower frequencies. In contrast, Geurts and Wouters' modified processing strategy employed triangular filters, designed to have a sharply peaked frequency response, and, ideally, a maximum width of one octave. The aim was that the first harmonic would be resolved in two adjacent filters, and changes in its frequency would be closely reflected in the relative outputs of the two filters. McDermott and McKay (1994) found that when pulse trains were presented simultaneously to two adjacent electrodes, implant users perceived a single pitch which varied according to the relative difference in levels between the pulses on the two electrodes. On this basis it was expected that the modified filtering would result in an effective place cue to changes in voice pitch. This expectation was supported by the fact that the four users of the LAURA cochlear implant tested were generally better at discriminating differences in the *F*0 of synthesized monopthongs with the modified, rather than the typical, filtering. However, it remains unclear whether this approach will still be effective with more natural speech stimuli. Natural speech typically contains dynamic variations in spectral shape, resulting in changes in the distribution of energy in different channels which may conflict with and obscure the place cue to *F*0.

Other attempts to improve intonation perception in cochlear implant systems have focused on enhancing the temporal pitch cues derived from envelope modulations corresponding to *F*0. Geurts and Wouters (2001) implemented a modified version of CIS processing that featured a number of changes from standard strategies, including the subtraction of the output of a 50 Hz low-pass filter from that of the standard 400 Hz low-pass filter, resulting in a larger modulation depth for *F*0-related fluctuations. However, they did not find significant improvements in the ability of four users of the LAURA cochlear implant to discriminate changes in the *F*0 of synthesized monophthongs with the new strategy compared to standard CIS processing.

Green, Faulkner, and Rosen (2004) adopted a different approach to the enhancement of temporal pitch cues, based on the idea that *F*0 could be extracted from voiced segments of the speech input and reintroduced in a simplified form. In a modified processing strategy amplitude envelopes were effectively split into two separate components. One consisted of slow rate information (<32 Hz) conveying the dynamic changes in spectral shape that are crucial for speech; the second presented *F*0 information in the form of a simplified synthesized wave form, allowing complete control over the form in which *F*0-related information was presented. In addition to maximizing the depth of the *F*0 modulation and allowing the same wave form to be applied to each channel, a key feature of such an approach is that the shape of the modulation wave form can be optimized so as to maximize the salience of temporal pitch cues.

In the experiments conducted by Green, Faulkner, and Rosen (2004), *F*0-related modulation was presented in the form of a sawtooth wave form, on the assumption that such a "temporally sharpened" modulation envelope, with a rapid onset in each period, would lead to more consistent interpulse intervals in the neural firing pattern, and therefore to more salient temporal pitch cues. Implant users, and normally hearing subjects listening to noise-excited vocoder simulations, were required to label the direction of pitch movement of processed synthetic diphthong glides. In both cases there was a significant advantage for modified compared to standard CIS processing, indicating that the modified processing scheme was successful in enhancing the salience of temporal pitch cues. It also appeared that, at least in some cases, performance was better when the sawtooth modulation was subject to additional temporal sharpening, such that the fall from peak to zero occurred in the first half of each period, and the second half of each period remained at zero.

Conditions in which slow-rate spectral dynamics were eliminated, so that within-channel amplitude changes reflected only *F*0, showed that the effectiveness of temporal cues to pitch was hindered by the variations in spectral envelope caused by the changing formant structure of the diphthongal vowel glides. This finding emphasizes the fact that cues that may provide useful pitch information in the context of certain simplified stimuli may not be robust in the presence of natural speech. It is also possible that the *F*0 modulation patterns associated with synthetic speech stimuli are less temporally complex than those of natural speech, which might, for example, feature secondary peaks within modulation periods. If so the simplified modulation wave form provided by Green, Faulkner, and Rosen's (2004) modified strategy may have additional benefits relative to traditional CIS processing.

The current study carried out further comparisons of the modified strategy with standard CIS processing, which again were performed both in implant users and in acoustic simulations. In this study only the modified sawtooth wave form, the additional temporal sharpening of which appeared beneficial in the previous study, was used to convey $F0$-related modulation. Pitch perception abilities were assessed using natural speech stimuli in tests of the ability to distinguish between questions and statements, a task which reflects an everyday use of intonation information. As a first step toward ensuring that benefits in pitch perception are not achieved at the expense of other linguistic information, vowel recognition scores were obtained. In addition, implant users were tested on their ability to discriminate differences in formant frequencies, as there is evidence that this ability is a primary determinant of implant users' vowel perception (Harnsberger *et al.*, 2001).

II. ACOUSTIC SIMULATION EXPERIMENTS

A. Subjects

Five undergraduate students aged between 22 and 25 who were native speakers of British English took part in both acoustic simulation experiments. They were paid £5 per hour for their participation.

B. Speech processing

1. Standard CIS

Noise-excited vocoder processing was carried out as in Green, Faulkner, and Rosen (2004). Simulation of standard CIS processing comprised the following sequence of steps: analysis bandpass filtering (sixth-order Butterworth) to divide the spectrum into eight frequency bands; full-wave rectification and low-pass filtering (second-order Butterworth) with a cutoff frequency of 400 Hz to extract the amplitude envelope for each band; modulation of independent noise carriers by these envelopes; output filtering matching the initial analysis filtering; summation across channels. The cutoff frequencies of the analysis filters, which were based on the Clarion S-Series Pulsatile Table with extended low- and high-frequency settings (Clarion Device Fitting Manual, 2001), were 250, 500, 730, 1015, 1450, 2000, 2600, 3800, and 6800 Hz.

2. Enhanced $F0$ (Sawsharp)

Processing in the Sawsharp condition differed from standard CIS processing in two respects. First, the cutoff frequency of the low-pass filters used in envelope extraction was 32 Hz rather than 400 Hz, eliminating $F0$-related fluctuations but passing the low-rate envelope information essential for speech perception. Second, during voiced sections of the speech input, temporal pitch cues were introduced by modulating the noise carrier in each channel with a sharpened sawtooth wave form with periodicity matching that of the speech input. The detection of voicing and the periodicity of the voiced segments were determined from laryngograph recordings. The modulation depth of the sharpened sawtooth wave form was always 100%.

C. Stimuli and procedure

1. Question/Statement identification

Thirty sentences were used. They included eight sentences used by Lieberman and Michaels (1962) in their investigation of the role of pitch in conveying emotional content in speech; thirteen sentences from the Question/Statement subtest of the Minimal Auditory Capabilities Battery (Owens *et al.*, 1981); and nine other simple declarative sentences (e.g., "They're playing in the garden;" "She's reading a newspaper"). Simultaneous audio and laryngograph recordings were made in an anechoic room of each of the sentences being read three times as a statement (falling pitch contour), and three times as a question (rising pitch contour), by one male and one female native speaker of Southern British English. The range of $F0$ values was approximately 100–220 Hz for the male speaker, and 120–360 Hz for the female speaker. Example $F0$ contours are shown in Fig. 1.

Blocks of trials contained 10 of the 30 sentences spoken as both question and statement by both speakers. Both processing conditions were included within each block resulting in a total of 80 trials per block. A total of nine blocks were presented in random order, incorporating each of the three versions of each utterance from each speaker. Stimuli were presented diotically through Sennheiser HD 414 headphones at a comfortable listening level (peak of 85–90 dB SPL measured over an 80 ms window). On each trial subjects heard a single sentence and were required to identify it as either "question" or "statement." They responded via computer mouse; visual feedback was provided. Before each block of trials subjects were able to listen to a selection of the sentences to be presented in that block, visually labeled as "question" or "statement."

2. Vowel recognition

Stimuli were selected from recordings of monophthongs in a /bVd/ context spoken by one male and one female native speaker of Southern British English. Simultaneous audio and laryngograph recordings had again been made in an anechoic room, but the speakers were different from those who read the Question/Statement sentences. The selected stimuli comprised two recordings from each speaker of the words "bead," "bird," "board," "bard," "booed," "bid," "bed," "bad," "bud," "bared," "bode," and "beard."

Processing condition was varied across blocks of 48 trials (12 vowels × 2 speakers × 2 exemplars). Four blocks were completed for each processing condition with the first counted as practice. Stimuli were presented as in the Question/Statement task. Responses were made by clicking on the desired word on a computer screen showing all twelve possible answers. Feedback was provided by highlighting the correct word after each response. Before each block listeners could listen to as many examples of each word as they wished by clicking on that word on the screen.

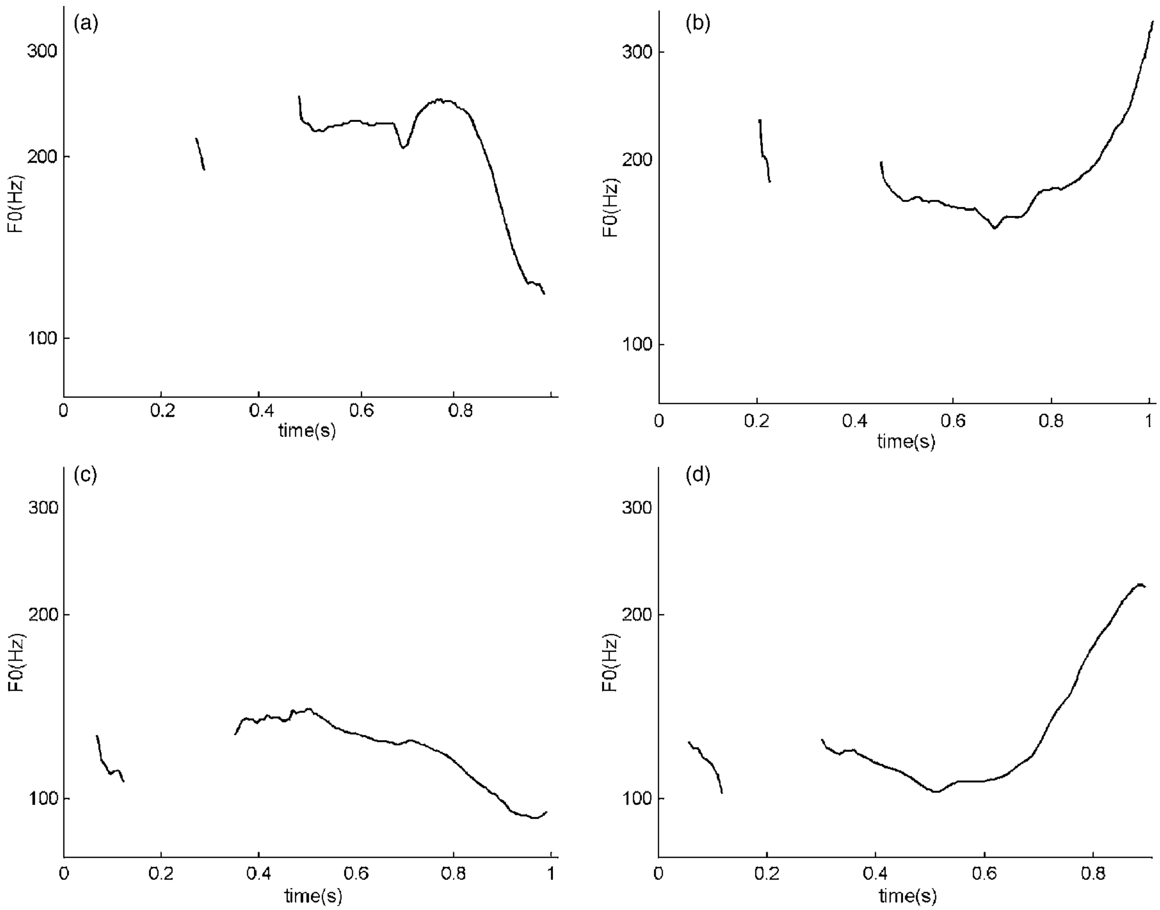

FIG. 1. Fundamental frequency contours for four different versions of the sentence "it's down there." (a) Female speaker, statement; (b) female speaker, question; (c) male speaker, statement; (d) male speaker, question.

D. Results and discussion

1. Question/Statement identification

In Fig. 2 Question/Statement identification performance with Sawsharp processing is plotted against performance with CIS processing, for each subject and speaker. Performance was consistently higher with Sawsharp than with CIS processing (all points are on or above the diagonal). With the exception of subject S5 with CIS processing, performance was better for the female than for the male speaker (open symbols are to the right of, and above, their closed counterparts). Mean performance with CIS processing was 78.2% with the male speaker and 81.8% with the female, while with Sawsharp processing the means were 81.8% with the male speaker and 87.3% with the female. A repeated measures ANOVA with factors of type of processing and speaker gender showed a significant effect of processing $[F(1,4)=20.85, p=0.010]$, but neither the effect of speaker gender $[F(1,4)=1.85, p=0.245]$, nor the interaction $[F(1,8)=0.43, p=0.547]$ was significant. In order to assess possible biases in responding, signal detection theory techniques were used to calculate sensitivity (d') and response criterion (c) across speaker gender for each listener and type of processing. Values of d' showed the same pattern as percent correct. Values of c were almost exclusively positive, indicating a bias toward "statement" responses, but there was no clear change of response criterion according to processing condition.

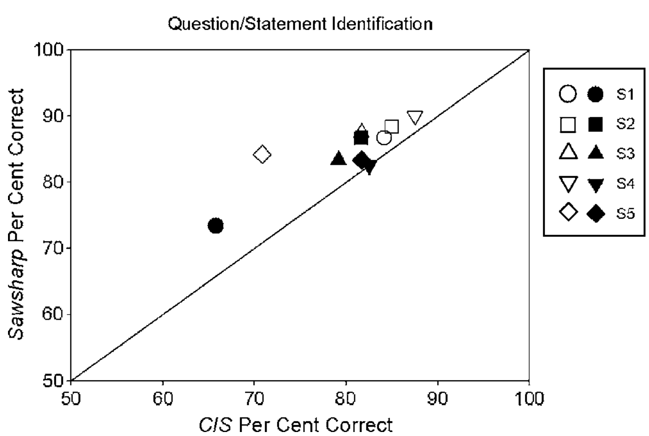

FIG. 2. Percent correct Question/Statement identification for each normally hearing listener for the male and female speakers (closed and open symbols, respectively). Performance in the Sawsharp condition is plotted against that in the CIS condition. The diagonal line represents equal performance in the two conditions.

The increased ability to identify sentences as questions or statements with the modified strategy compared to standard CIS processing presumably reflects the greater transmission of intonation information with modified processing. However, it should be acknowledged that other cues, such as differences in amplitude contour across questions and statements, may have been available. Such cues probably contrib-

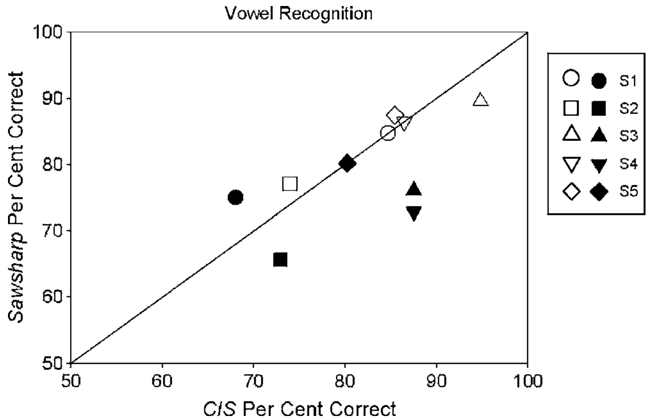

FIG. 3. Percent correct vowel recognition for each normally hearing listener for the male and female speakers (closed and open symbols, respectively). Performance in the Sawsharp condition is plotted against that in the CIS condition. The diagonal line represents equal performance in the two conditions.

uted to the relatively high scores obtained even in the standard CIS condition. It is noticeable that, despite the higher voice pitch of the female speaker, and the fact that temporal envelope cues to pitch decline in utility with increasing $F0$, performance is in fact slightly better for the female speaker than the male. One possible explanation is that amplitude contour differences may have been more pronounced for the female speaker. In addition, the $F0$ contours shown in Fig. 1 suggest that the female speaker displayed greater variation in $F0$, which would help to overcome the lesser utility of temporal envelope cues to pitch at higher $F0$s. Regardless of such issues, the significant advantage for modified processing strongly suggests that this strategy enhances the salience of temporal cues to voice pitch, as had been found previously with synthetic speech materials (Green, Faulkner, and Rosen, 2004).

2. Vowel recognition

Averaged across speakers, vowel recognition with the different types of processing was very similar. Performance was generally better with the female than with the male speaker (Fig. 3). Averaged across listeners, performance was identical across processing conditions for the female speaker (85.1%), but was slightly poorer with Sawsharp processing for the male speaker (74.0% compared to 79.2%). A repeated measures ANOVA showed a significant effect of speaker gender $[F(1,4)=35.08, p=0.004]$, but neither the effect of processing condition $[F(1,4)=1.33, p=0.314]$, nor the interaction $[F(1,4)=2.04, p=0.227]$ were significant. Thus, in acoustic simulations, the modified processing strategy delivers enhanced temporal pitch cues without harming the transmission of spectral information necessary for vowel recognition. The significant effect of speaker gender may result from the higher formant frequencies of the female speaker, which, in conjunction with the particular analysis filter cutoff frequencies used, may have produced distributions of energy across channels that were more different for similar vowels for the female than the male speaker.

III. IMPLANT USERS

A. Subjects and equipment

A total of nine post-lingually deafened adult users of the eight-channel Clarion 1.2 cochlear implant system participated. At least seven subjects took part in each test. Summary information is contained in Table I. Eight of the subjects had taken part in the vowel glide labeling experiments reported in Green, Faulkner, and Rosen (2004). As in that study, experiments were controlled by a PC connected to a Clarion Research Interface (CRI) system (Wygonski *et al.*, 1999), allowing direct control over the stimulus patterns presented to the electrode array. Stimulation consisted of con-

TABLE I. Demographic information.

Subject	Age	Gender	Cause of deafness	Age at onset	Age at implantation	Implant type	Strategy[a]	Mean threshold (clinical units)	Mean MCL (clinical units)
C1	73	M	Otosclerosis	69	70	Enhanced bipolar	SAS	93	329
C2	75	F	Unknown	69	71	Enhanced bipolar	MPS	97	397
C3	71	M	Unknown	40	68	Enhanced bipolar	MPS	60	633
C4	72	M	Skull fracture	64	69	Hifocus 1.2	SAS	49	228
C5	61	F	Unknown	51	59	Hifocus 1.2	MPS	40	138
C6	45	F	Unknown	30	41	Hifocus 1.2	MPS	26	499
C7	65	F	Unknown	45	62	Hifocus 1.2	MPS	85	174
C8	56	F	Sensorineural	42	51	Enhanced bipolar	CIS	100	614
C9	57	F	Sensorineural	14	55	Hifocus 1.2	SAS	21	133

[a]MPS is a variant of the CIS strategy in which pairs of channels are stimulated simultaneously.

tinuously interleaved, monopolar, biphasic pulses with a duration of 76.9 μs per phase. The carrier rate was 812.5 pulses per second per electrode (ppse) and electrodes were activated sequentially in apical-to-basal order.

B. Speech processing

A full description of the processing methods appears in Green, Faulkner, and Rosen (2004); only brief details are given here. Analysis filtering and envelope extraction were implemented as in the acoustic simulations for both modified and standard CIS processing. Envelopes were resampled to a rate of 6500 Hz, consistent with the overall pulse rate. In the CIS condition pulse levels were determined by logarithmic compression and mapping of a 50 dB range of envelope sample values onto the dynamic range of the particular subject and channel. In the Sawsharp condition the starting point of each successive period of $F0$ during voiced speech segments was determined from the laryngograph recording. During these segments, the carrier for each channel consisted of a pulse train 100% amplitude modulated by a sharpened sawtooth wave form; when voicing was absent the carriers were simply unmodulated pulse trains. The final stage of processing consisted of modulation of the pulse carriers by low-rate amplitude envelopes with compression and mapping carried out as in CIS processing.

C. Stimuli and procedure

Implant users carried out the same two tasks that were used in the acoustic simulations, with the exception that vowel recognition was tested using a smaller set of stimuli. This reflected the fact that the implant users' performance was substantially poorer than that of normally hearing listeners. Seven implant users were tested with nine, rather than twelve, vowels, with the words "bared," "bode," and "beard" omitted. The remaining two (C3 and C9) were tested with just five vowels: "bad," "bard," "bead," "bird," and "booed."

Implant users also performed a formant frequency discrimination task. Stimuli consisted of processed versions of a continuum of synthetic two-formant vowel sounds. These were created using an implementation of the KLSYN88 Klatt synthesizer in cascade mode with a 20 kHz sample rate and parameters specified every 5 ms. The first formant frequency was fixed at 500 Hz, while the second ($F2$) was varied between 1500 and 3000 Hz in 30 equal logarithmic steps. $F0$ declined logarithmically from 139 to 92 Hz over the 500 ms duration. A three-interval, 2AFC task with an adaptive two-down, one-up procedure was used. The first interval always had $F2$ at 1500 Hz. The second and third intervals contained, in random order, one stimulus with the same $F2$ and one with a higher $F2$. The time gap between intervals was 500 ms. The listener's task was to identify the interval containing the higher $F2$. In order to minimize overall loudness cues, the levels of stimuli within a trial were varied pseudo-randomly. Four different levels were employed, differing in 2 dB steps. The combination of levels of the three stimuli in a particular trial was selected at random from a subset of 8 of the total of 64 possible level combinations.

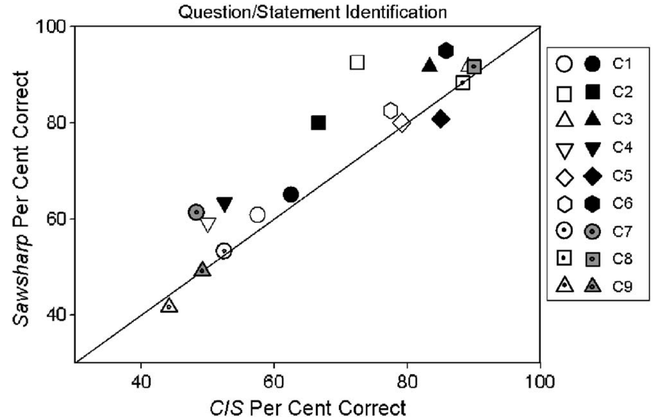

FIG. 4. Percent correct Question/Statement identification for each implant user for the male and female speakers (closed and open symbols, respectively). Performance in the Sawsharp condition is plotted against that in the CIS condition. The diagonal line represents equal performance in the two conditions.

In the first trial of each run the stimulus to be identified had $F2$ at 3000 Hz. Over the first three reversals the step size was gradually reduced from 5 (out of 30) steps to 1. A threshold was calculated for each run from the average of eight subsequent reversals. Reported threshold values for each processing condition are the average of either five or six runs.

D. Results and discussion

1. Question/Statement identification

For both speakers, mean performance was around 5% better with Sawsharp than with CIS processing (75.4% compared to 69.3% for the male speaker and 72.2% compared to 67.9% for the female). Four subjects (C1, C5, C8, and C9) showed little difference between the two processing conditions (Fig. 4). Ceiling effects may have occurred for C8, but not in the other cases. The other five subjects showed an advantage for Sawsharp processing of between approximately 5% and 15%. A repeated measures ANOVA showed a significant effect of type of processing $[F(1,8)=7.17, p=0.028]$, while neither the effect of speaker gender $[F(1,8)=1.50, p=0.256]$, nor the interaction $[F(1,8)=0.85, p=0.383]$ were significant.

Values of response criterion (c) also displayed considerable variability but were generally positive, reflecting a bias toward "statement" responses. In contrast to the acoustic simulation data, there was a consistent different between Sawsharp and CIS processing, with the former resulting in an increased proportion of "question" responses. However, a repeated measures t-test comparing d' values showed that the advantage for Sawsharp processing was not dependent upon this change in response bias $[t(8)=2.85, p=0.022]$.

Both the absence of an effect of speaker gender and the significant advantage for the modified processing over standard CIS are consistent with the data obtained in the noise-excited vocoder acoustic simulation. As was also the case in the glide labeling task used by Green, Faulkner, and Rosen (2004), mean performance by implant users was poorer than that of normally hearing listeners in acoustic simulations, but the better implant users achieved scores similar to those ob-

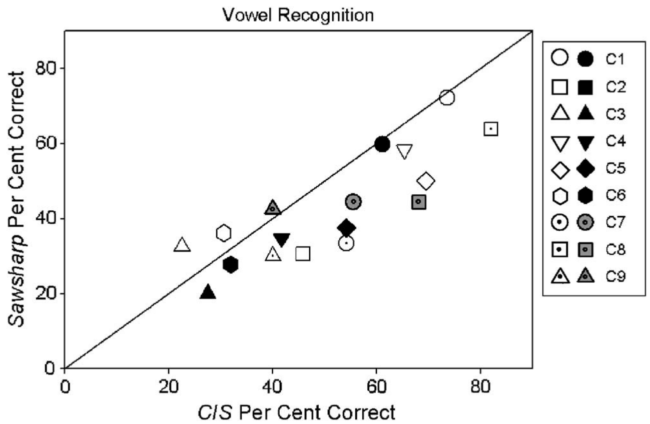

FIG. 5. Percent correct vowel recognition for each implant user for the male and female speakers (closed and open symbols, respectively). Performance in the Sawsharp condition is plotted against that in the CIS condition. The diagonal line represents equal performance in the two conditions.

tained in the simulations. This provides further evidence that the temporal cues to voice pitch available in noise-excited vocoder simulations are broadly similar to those available with CIS-like processing. Most important, these data, obtained with natural speech stimuli in a task reflecting an everyday use of intonation information, indicate that the modified processing scheme, by providing enhanced $F0$-related modulation, leads to better voice pitch perception.

2. Vowel recognition

In contrast to the acoustic simulation data, vowel recognition (Fig. 5) was substantially affected by processing condition. Averaged across listeners, performance for the male speaker was 47.3% with CIS processing compared to 38.0% with Sawsharp processing, while for the female speaker the respective means were 53.7% and 45.2%. Averaged across speakers, performance varied little across processing condition for four subjects (C1, C3, C6, and C9), but was substantially poorer with Sawsharp processing for the remaining five subjects. A repeated measures ANOVA showed a significant effect of processing condition $[F(1,8)=9.57, p=0.015]$, but neither the effect of speaker gender $[F(1,8)=3.85, p=0.085]$, nor the interaction $[F(1,8)=0.08, p=0.786]$ were significant.

To further investigate this deficit, sequential information analysis (SINFA, Wang and Bilger, 1973) was carried out. Only data from the seven subjects who were tested on all nine vowels were included in these analyses. The features on which vowels were classified were *duration, open*, and *front* (Table II). Assignment of vowels to different feature categories followed the standard IPA classification, with one exception. Although "bad" is typically classified as a short vowel, in the stimuli used here it was the longest vowel for the male speaker and the third longest for the female speaker, and was therefore classified as long. SINFA was performed for each combination of speaker and processing condition, both on individual implant users' confusion matrices and on pooled data (Fig. 6).[11]

The analysis of the pooled data (Table III) shows that, for both speakers, and for all three features, less information

TABLE II. Classification of the vowel features used in SINFA: o=open; om=open-mid; cm=closed-mid; c=closed; f=front; ce=central; b=back; l=long; s=short. Standard IPA classifications were used with one exception. Although "bad" is typically classified as a short vowel, in the stimuli used here it was the longest vowel for the male speaker and the third longest for the female speaker, and was thus classified as long.

	Bad æ	Bard ɑ	Bead i	Bed e	Bid ɪ	Bird ɜ	Board ɔ	Booed u	Bud ʌ
Open	o	o	c	om	c	cm	om	c	om
Front	f	b	f	f	f	ce	b	b	b
Duration	l	l	l	s	s	l	l	l	s

is transmitted with Sawsharp than with standard CIS processing. The results of the analysis of individual data (Fig. 7) are largely consistent with the pooled data, though there is much individual variability. The decrease in transmitted information for the *open* and *front* features with Sawsharp processing suggests that the modified processing strategy, while providing enhanced pitch information, also has a harmful effect on the perception of spectral information essential for speech understanding.

Surprisingly, the proportion of duration information transmitted also declined with the modified processing. One relevant factor may be that the male speaker's stimuli were consistently longer than those of the female (mean 0.69 s, compared to 0.53 s). Because stimuli from both speakers were presented in the same block of trials, it is plausible that the ability to identify the speaker's gender limits the extent to which it is possible to make full use of duration information. Notwithstanding the improved representation of $F0$ information with modified processing, it is likely that the primary cue distinguishing male and female voices is the different distribution of energy across frequency channels due to differences in formant frequencies. Thus, to the extent that the ability to make use of duration cues depended on being able to distinguish between the two speakers, a decline in the transmission of spectral information might result in a decline in the availability of duration information. An alternative possibility concerns the reduction in the cutoff frequency of

***CIS* Male Speaker**

	æ	ɑ	i	e	ɪ	ɜ	ɔ	u	ʌ
æ	25	19	0	0	0	10	1	1	0
ɑ	6	43	0	0	0	3	4	0	0
i	6	1	29	1	7	3	4	5	0
e	5	0	0	32	2	1	0	0	16
ɪ	5	1	1	32	11	1	2	0	3
ɜ	12	10	0	0	0	28	4	2	0
ɔ	4	17	0	0	0	5	28	2	0
u	2	15	0	0	0	5	13	21	0
ʌ	3	0	0	11	0	1	0	0	41

***CIS* Female speaker**

	æ	ɑ	i	e	ɪ	ɜ	ɔ	u	ʌ
æ	27	6	1	1	1	11	1	0	8
ɑ	8	30	0	0	0	11	7	0	0
i	3	0	26	8	4	6	4	5	0
e	1	0	0	26	13	0	0	0	16
ɪ	0	0	0	7	42	0	0	0	7
ɜ	3	5	0	8	1	32	1	3	3
ɔ	0	0	0	0	0	1	52	3	0
u	1	2	0	1	0	19	12	19	2
ʌ	0	0	0	3	4	0	0	0	49

***Sawsharp* Male Speaker**

	æ	ɑ	i	e	ɪ	ɜ	ɔ	u	ʌ
æ	14	22	0	0	0	16	2	2	0
ɑ	6	32	0	0	0	13	4	1	0
i	9	2	24	4	1	4	2	10	0
e	3	0	0	29	3	4	0	3	14
ɪ	5	1	1	27	8	4	0	3	7
ɜ	19	16	0	0	0	14	6	1	0
ɔ	4	15	1	0	0	9	24	2	1
u	4	12	1	0	0	7	15	17	0
ʌ	6	0	0	7	1	2	0	1	39

***Sawsharp* Female Speaker**

	æ	ɑ	i	e	ɪ	ɜ	ɔ	u	ʌ
æ	6	9	1	2	0	21	3	8	6
ɑ	4	21	0	0	0	6	12	10	3
i	1	0	25	3	1	4	7	13	2
e	2	0	3	21	9	1	0	4	16
ɪ	0	0	1	6	41	0	0	3	5
ɜ	6	1	0	1	1	24	10	10	3
ɔ	0	1	1	0	0	4	45	3	2
u	1	1	1	2	0	8	17	24	2
ʌ	1	0	0	2	9	0	0	3	41

FIG. 6. Confusion matrices for each combination of processing type and speaker gender, representing pooled data from the seven subjects who were tested on all nine vowels.

TABLE III. Proportion of information transmitted based on SINFA performed on pooled data from the seven implant users tested with all nine vowels.

	Duration	Open	Front
CIS M	0.663	0.221	0.241
CIS F	0.678	0.354	0.283
Sawsharp M	0.578	0.139	0.146
Sawsharp F	0.541	0.208	0.155

the envelope filter from 400 to 32 Hz in the modified processing condition. Although we would not have expected this reduction to have significantly affected access to vowel duration differences, it is perhaps conceivable that poorer definition of the attack and decay portions of the envelope with the lower cutoff frequency could have contributed to the decline in transmission of duration information.

In order to assess any possible relationship between decrements in vowel recognition and enhancements in pitch perception, a Pearson's correlation coefficient was calculated for the differences between the processing conditions in vowel recognition and Question/Statement performance (averaged across speakers). The results $(r=-0.02, p=0.96)$ showed clearly that there was no such relationship. Thus, it is not the case that a greater benefit in pitch perception from modified processing was associated with a larger decrement in vowel recognition.

However, although there are not enough subjects to draw any strong conclusions, it does appear that the strategy normally used by the subjects influences the decrement in vowel recognition with modified processing. The decrement for the single CIS user was nearly 21%, while for the five users of the multiple pulsatile sampler (MPS) strategy the mean decrement was 9.5%, and for the three SAS users it was 4%. The MPS strategy is similar to CIS in that it presents interleaved stimulation, although in this case the interleaving is only partial, with two separate channels receiving pulses simultaneously. If, on this basis, users of MPS are considered to be more familiar with CIS processing than SAS users, then it would appear that the greater the degree of familiarity with CIS processing then the larger the advantage in vowel recognition for CIS than for Sawsharp processing.

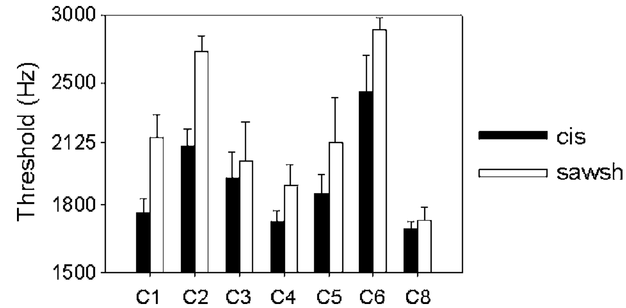

FIG. 8. Implant users' mean threshold $F2$ values for discrimination from a 1500 Hz standard. Error bars show one standard deviation.

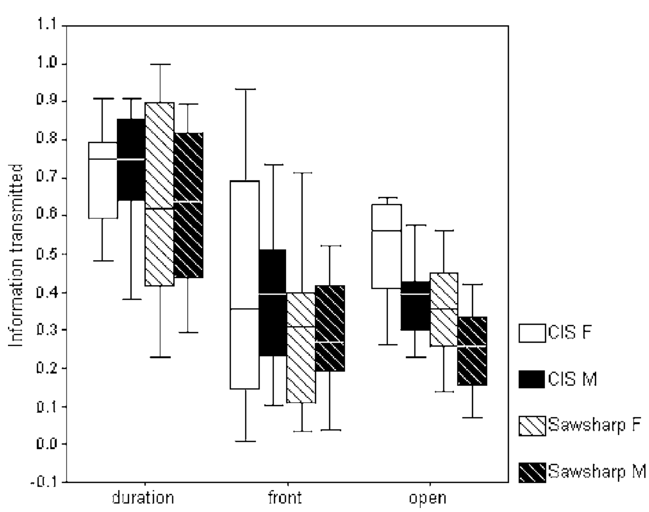

FIG. 7. Boxplot showing results of SINFA performed on confusion matrices from each of the seven implant users who were tested with all nine vowels.

3. Formant discrimination

Thresholds for discriminating a difference in $F2$ were higher with Sawsharp than with CIS processing for all seven subjects, although the difference in thresholds for the two types of processing varied substantially (Fig. 8). A repeated measures t-test showed that the effect of type of processing was significant $[t(6)=3.77, p=0.009]$. This provides further evidence suggesting a detrimental impact of the modified processing strategy on the perception of spectral information necessary for speech understanding.

Pearson correlation coefficients were calculated in order to assess the relationship between formant discrimination performance (expressed logarithmically) and vowel recognition (averaged across speakers). With CIS processing, the correlation just missed significance $(r=-0.723, p=0.067)$. With Sawsharp processing the relationship was not significant $(r=-0.505, p=0.248)$. If the data from subject C3, whose vowel recognition was tested with only five vowels, are omitted then substantially larger correlations are obtained. The correlation was significant with CIS processing $(r=-0.901, p=0.014)$ and close to significance with Sawsharp processing $(r=-0.724, p=0.104)$. Omitting the data of one further outlier (C1) in the Sawsharp condition resulted in a highly significant correlation $(r=-0.983, p=0.003)$. Clearly, considerably more subjects would be required to accurately determine the extent of the relationship between the two measures, though this limited sample does suggest a reasonably strong relationship.

However, there appears to be virtually no relationship between the effects of processing condition in the two different tasks. For example, subject C8 showed the largest decrement for Sawsharp processing in vowel recognition, but had the smallest difference in $F2$ discrimination thresholds. Conversely, for C1, $F2$ discrimination was much poorer with Sawsharp processing, but there was only a very small difference in vowel recognition. Consequently, the correlation coefficient for the percentage difference in vowel recognition and the difference in $F2$ discrimination threshold was minimal $(r=0.091, p=0.846)$.

IV. GENERAL DISCUSSION

As was the case with synthetic stimuli in Green, Faulkner, and Rosen (2004), relatively small, but significant, benefits in voice pitch perception for modified processing of natural speech were observed both in implant users and in acoustic simulations. However, for implant users there was a significant deficit in vowel recognition with the modified strategy. It appears that the modified processing strategy resulted in poorer transmission of spectral information relative to standard CIS processing, although it is not easy to identify specific causes of the decrement in vowel recognition. Of course, it is possible that different factors may contribute to differences between processing conditions for different implant users.

One major element of the modified scheme is that it relies upon slow-rate (<32 Hz) modulations to convey the dynamic spectral variation essential for speech understanding. There is considerable evidence to suggest that speech perception, as assessed by standard speech tasks such as vowel and consonant recognition, depends only upon low frequency temporal envelope information (e.g., Drullman, Festen, and Plomp, 1994; Fu and Shannon, 2000; Shannon *et al.*, 1995; Van Tasell *et al.*, 1992). In conjunction with the absence of an effect of processing condition in the acoustic simulation data, this suggests that while the reduction of the envelope filter cutoff frequency from 400 to 32 Hz in the modified processing condition may have some impact, it is unlikely to be primarily responsible for the decrement in vowel recognition observed in implant users.

One possible explanation of the deficit could be a form of conflict between temporal and spectral information. There is evidence that spectral and temporal information interact in the determination of implant users' perception of pitch (Green, Faulkner, and Rosen, 2004; Zeng, 2002). This suggests the possibility that the enhanced temporal pitch information provided by the modified processing scheme may interfere with implant users' perceptual processing of the spectral information that is encoded in between-channel differences in level. However, the absence of any relationship between benefits in pitch perception and deficits in vowel recognition would appear to count against such an explanation.

It may be important that in the Sawsharp instantiation of the modified processing strategy implemented here, the level of the pulse carrier for voiced speech declines from its maximum value to threshold over the first half of each period, and remains at threshold for the second half of each period. Thus, with this modulation wave form shape, only a small part of the voicing cycle is carrying spectral information. We used the Sawsharp wave form here because in our previous study (Green, Faulkner, and Rosen, 2004), it appeared to be close to optimal for conveying temporal cues to voice pitch. It was hoped that it would also adequately convey spectral information. Although the present findings suggest that this is not the case, it is important to bear in mind that subjects had very little experience of the Sawsharp processing strategy prior to testing. Particularly in light of the apparent relationship between the extent of the decrement in vowel recognition and the implant user's normal processing strategy, it is conceivable that, given sufficient experience, implant users might adjust to the different way in which spectral information is presented with Sawsharp processing, thereby eliminating the deficit relative to standard CIS processing.

Even if extra experience with Sawsharp processing were not sufficient to restore optimal vowel recognition performance, there are a number of potential changes to the modified processing strategy that may improve matters. One possibility concerns the shape of the synthesized wave form carrying $F0$-related modulation. While Green, Faulkner, and Rosen (2004) found that performance in their glide labeling task was best in the Sawsharp condition, there was only a small advantage over a condition that used a standard sawtooth wave form, in which the decline from maximum to minimum level occurred over the whole period. If the decrement in vowel perception with Sawsharp processing is related to the small proportion of the voicing cycle that is available to convey spectral information, then using a standard sawtooth wave form may overcome this deficit.

The modulation depth of the simplified $F0$ wave form could also be manipulated. Fu (2002) showed that cochlear implant users' modulation detection thresholds decrease with increasing stimulation level, although the pattern of this decrease varies markedly across individuals. This suggests that an approach in which the depth of the $F0$-related modulation was varied according to stimulation level, with the pattern of the variation determined on an individual basis, may help to preserve the benefits to voice pitch perception provided by the modified processing strategy while minimizing the harmful impact on vowel perception.

Another possibility concerns the number of channels to which the clarified $F0$-related modulation is applied. Based on evidence of better discrimination of the modulation frequency of sinusoidally amplitude modulated pulse trains with three adjacent channels stimulated concurrently, compared to any one channel alone (Geurts and Wouters, 2001), the modified processing scheme was implemented with the simplified $F0$ wave form applied to all eight channels. However, it is conceivable that applying the clarified $F0$ modulation to a subset of channels might provide enhanced pitch cues, while affording better transmission of spectral information.

It should also be borne in mind that the present research has been carried out in users of implant systems that have a pulse rate of approximately 800 ppse. Newer implant systems are typically capable of much faster pulse rates (e.g., 5000 ppse) which may provide greater scope for using the kind of approach adopted here to provide enhanced temporal cues to voice pitch.

ACKNOWLEDGMENTS

The authors thank all participants and are grateful to the cochlear implant team at the Royal National Throat Nose and Ear Hospital for help in recruiting implant users. Ken Grant, Fan-Gang Zeng, and Monita Chatterjee made helpful comments on an earlier version of this manuscript. This research was supported by the Royal National Institute for Deaf People (UK).

[1]Typically in SINFA, the feature for which the proportion of transmitted information is highest is held constant for subsequent iterations. In order to allow comparisons across different conditions, in the few cases where the order of analysis differed from the most common sequence (*duration, open, front*), analyses were repeated with this order and it is these results that are reported.

Au, D. K. K. (**2003**). "Effects of stimulation rates on Cantonese lexical tone perception by cochlear implant users in Hong Kong," Clin. Otolaryngol. **28**, 533–538.

Barry, J. G., Blamey, P. J., Martin, L. F. A., Lee, K. Y.-S., Tang, T., Ming, Y. Y., and Van Hasselt, C. A. (**2002**). "Tone discrimination in Cantonese-speaking children using a cochlear implant," Clin. Linguist. Phonet. **16**, 79–99.

Burns, E. M., and Viemeister, N. F. (**1976**). "Nonspectral pitch," J. Acoust. Soc. Am. **60**, 863–869.

Burns, E. M., and Viemeister, N. F. (**1981**). "Played-again SAM: Further observations on the pitch of amplitude-modulated noise," J. Acoust. Soc. Am. **70**, 1655–1660.

Busby, P. A., Tong, Y. C., and Clark, G. M. (**1993**). "The perception of temporal modulations by cochlear implant patients," J. Acoust. Soc. Am. **94**, 124–131.

Cazals, Y., Pelizzone, M., Saudan, O., and Boex, C. (**1994**). "Low-pass filtering in amplitude modulation detection associated with vowel and consonant identification in subjects with cochlear implants," J. Acoust. Soc. Am. **96**, 2048–2054.

Ciocca, V., Francis, A. L., Aisha, R., and Wong, L. (**2002**). "The perception of Cantonese lexical tones by early-deafened cochlear implantees," J. Acoust. Soc. Am. **111**, 2250–2256.

Donaldson, G., and Viemeister, N. (**2002**). "TMTFs in cochlear implant users: The role of loudness clues," Assoc. Res. Otolaryngol. Abstr. **25**, 121.

Drullman, R., Festen, J. M., and Plomp, R. (**1994**). "Effect of temporal envelope smearing on speech reception," J. Acoust. Soc. Am. **95**, 1053–1064.

Formby, C. (**1985**). "Differential sensitivity to tonal frequency and to the rate of amplitude modulation of broadband noise by normally hearing listeners," J. Acoust. Soc. Am. **78**, 70–77.

Fourcin, A. J., Rosen, S. M., Moore, B. C. J., Douek, E. E., Clarke, G. P., Dodson, H., and Bannister, L. H. (**1979**). "External electrical stimulation of the cochlea: Clinical, psychophysical, speech-perceptual and histological findings," Br. J. Audiol. **13**, 85–107.

Fu, Q.-J. (**2002**). "Temporal processing and speech recognition in cochlear implant users," NeuroReport **13**, 1635–1639.

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.

Fu, Q.-J., Zeng, F.-G., Shannon, R. V., and Soli, S. D. (**1998**). "Importance of tonal envelope cues in Chinese speech recognition," J. Acoust. Soc. Am. **104**, 505–510.

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.

Geurts, L., and Wouters, J. (**2004**). "Better place-coding of fundamental frequency in cochlear implants," J. Acoust. Soc. Am. **115**, 844–852.

Grant, K., Summers, V., and Leek, M. (**1998**). "Modulation rate detection and discrimination by normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **104**, 1051–1060.

Green, T., Faulkner, A., and Rosen, S. (**2002**). "Spectral and temporal cues to pitch in noise-excited vocoder simulations of continuous-interleaved-sampling cochlear implants," J. Acoust. Soc. Am. **112**, 2155–2164.

Green, T., Faulkner, A., and Rosen, S. (**2004**). "Enhancing temporal cues to voice pitch in continuous interleaved sampling cochlear implants," J. Acoust. Soc. Am. **116**, 2298–2310.

Hanna, T. (**1992**). "Discrimination and identification of modulation rate using a noise carrier," J. Acoust. Soc. Am. **91**, 2122–2128.

Harnsberger, J. D., Svirsky, M. A., Kaiser, A. R., Pisoni, D. B., Wright, R., and Meyer, T. A. (**2001**). "Perceptual 'vowel spaces' of cochlear implant users: Implications for the study of auditory adaptation to spectral shift," J. Acoust. Soc. Am. **109**, 2135–2145.

Jusczyk, P. (**1997**). *The Discovery of Spoken Language* (MIT, Cambridge, MA).

Lan, N., Nie, K. B., Gao, S. K., and Zeng, F.-G. (**2004**). "A novel speech processing strategy incorporating tonal information for cochlear implants," IEEE Trans. Biomed. Eng. **51**, 752–760.

Lee, K. Y. S., van Hasselt, C. A., Chiu, S. N., and Cheung, D. M. C. (**2002**). "Cantonese tone perception ability of cochlear implant children in comparison with normal-hearing children," Int. J. Pediatr. Otorhinolaryngol. **63**, 137–147.

Lieberman, P., and Michaels, S. B. (**1962**). "Some aspects of fundamental frequency and envelope amplitude as related to the emotional content of speech," J. Acoust. Soc. Am. **34**, 922–927.

McDermott, H. J., and McKay, C. M. (**1994**). "Pitch ranking with non-simultaneous dual-electrode electrical stimulation of the cochlea," J. Acoust. Soc. Am. **96**, 155–162.

McDermott, H. J., and McKay, C. M. (**1997**). "Musical pitch perception with electrical stimulation of the cochlea," J. Acoust. Soc. Am. **101**, 1622–1631.

McKay, C. M., and McDermott, H. J. (**1993**). "Perceptual performance of subjects with cochlear implants using the Spectral Maxima Sound Processor (SMSP) and the Mini Speech Processor (MSP)," Ear Hear. **14**, 350–367.

McKay, C. M., McDermott, H. J., and Clark, G. M. (**1994**). "Pitch percepts associated with amplitude-modulated current pulse trains by cochlear implantees," J. Acoust. Soc. Am. **96**, 2664–2673.

McKay, C. M., McDermott, H. J., Vandali, A. E., and Clark, G. M. (**1992**). "A comparison of speech perception of cochlear implantees using the Spectral Maxima Sound Processor (SMSP) and the MSP (Multipeak) processor," Acta Oto-Laryngol. **112**, 752–761.

Owens, E., Kessler, D., Telleen, C. C., and Schubert, E. (**1981**). "The minimal auditory capabilities (MAC) battery," Hearing Aid J. **9**, 9–34.

Peng, S.-C., Tomblin, J. B., Cheung, H., Lin, Y.-S., and Wang, L.-S. (**2004**). "Perception and production of Mandarin tones in prelingually deaf children with cochlear implants," Ear Hear. **25**, 251–263.

Pijl, S., and Schwarz, D. W. F. (**1995**). "Melody recognition and musical interval perception by deaf subjects stimulated with electrical pulse trains through single cochlear implant electrodes," J. Acoust. Soc. Am. **98**, 886–895.

Pollack, I. (**1969**). "Periodicity pitch for white noise—fact or artefact," J. Acoust. Soc. Am. **45**, 237–238.

Rubinstein, J. T., Wilson, B. S., Finley, C. C., and Abbas, P. J. (**1999**). "Pseudospontaneous activity: Stochastic independence of auditory nerve fibers with electrical stimulation," Hear. Res. **127**, 108–118.

Seligman, P., and McDermott, H. (**1995**). "Architecture of the Spectra 22 speech processor," Ann. Otol. Rhinol. Laryngol. Suppl. **166**, 139–141.

Shannon, R. (**1992**). "Temporal modulation transfer functions in patients with cochlear implants," J. Acoust. Soc. Am. **91**, 2156–2164.

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

Skinner, M. W., Fourakis, M. S., Holden, T. A., Holden, L. K., and Demorest, M. E. (**1999**). "Identification of speech by cochlear implant recipients with the Multipeak (MPEAK) and Spectral Peak (SPEAK) speech coding strategies," Ear Hear. **20**, 443–460.

Tong, Y. C., Millar, J. B., Clark, G. M., Martin, L. F., Busby, P. A., and Patrick, J. F., (**1980**). "Psychophysical and speech perception studies on two multiple channel cochlear implant patients," J. Laryngol. Otol. **94**, 1241–1256.

Townshend, B., Cotter, N., Van Compernolle, D. V., and White, R. L. (**1987**). "Pitch perception by cochlear implant subjects," J. Acoust. Soc. Am. **82**, 106–115.

Vandali, A. E., Whitford, L. A., Plant, K. L., and Clark, G. M. (**2000**). "Speech perception as a function of electrical stimulation rate: Using the Nucleus 24 cochlear implant system," Ear Hear. **21**, 608–624.

Van Tasell, D. J., Greenfield, D. G., Logemann, J. J., and Nelson, D. A., (**1992**). "Temporal cues for consonant recognition: Training, talker generalization, and use in evaluation of cochlear implants," J. Acoust. Soc. Am. **92**, 1247–1257.

Wang, M. D. and Bilger, R. C. (**1973**). "Consonant confusions in noise: A study of perceptual features," J. Acoust. Soc. Am. **54**, 1248–1266.

Wei, C.-G., Cao, K., and Zeng, F.-G. (**2004**). "Mandarin tone recognition in Chinese cochlear-implant subjects," Hear. Res. **197**, 87–95.

Wei, W. I., Wong, R., Hui, Y., Au, D. K. K., Wong, B. Y. K., Ho, W. K., Tsang, A., Kung, P., and Chung, E. (**2000**). “Chinese tonal language rehabilitation following cochlear implantation in children,” Acta Oto-Laryngol. **120**, 218–221.

Wilson, B. (**1997**). “The future of cochlear implants,” Br. J. Audiol. **31**, 205–225.

Wilson, B., Finley, C., Lawson, D., Wolford, R., Eddington, D., and Rabinowitz, W. (**1991**). “Better speech recognition with cochlear implants,” Nature (London) **352**, 236–238.

Wilson, B., Zerbi, M., Finley, C., Lawson, D., and van den Honert, C. (1997). Eighth Quarterly Progress Report, 1 May through 31 July 1997. NIH Project N01-DC-5-2103: Speech Processors for Auditory Prostheses: Research Triangle Institute.

Wygonski, J. J., Lee, J., Faltys, M., Shannon, R. V., and Robert, M. (**1999**). “Configurable speech strategy implementation using the Clarion research interface,” 1999 Conference on Implantable Auditory Prostheses, Pacific Grove, CA.

Xu, L., Tsai, Y., and Pfingst, B. E. (**2002**). “Features of the stimulation affecting tonal-speech perception: Implications for cochlear prostheses,” J. Acoust. Soc. Am. **112**, 247–258.

Zeng, F.-G. (**2002**). “Temporal pitch in electric hearing,” Hear. Res. **174**, 101–106.