The perception of Cantonese lexical tones by early-deafened cochlear implantees

Valter Ciocca,[a)] Alexander L. Francis, Rani Aisha, and Lena Wong
Department of Speech & Hearing Sciences, University of Hong Kong, 5/F, Prince Philip Dental Hospital, 34 Hospital Road, Hong Kong SAR, China

(Received 25 July 2001; accepted for publication 26 February 2002)

This study investigated whether cochlear implant users can identify Cantonese lexical tones, which differ primarily in their $F0$ pattern. Seventeen early-deafened deaf children (age = 4 years, 6 months to 8 years, 11 months; postoperative period = 11–41 months) took part in the study. Sixteen children were fitted with the Nucleus 24 cochlear implant system; one child was fitted with a Nucleus 22 implant. Participants completed a 2AFC picture identification task in which they identified one of the six contrastive Cantonese tones produced on the monosyllabic target word /ji/. Each target stimulus represented a concrete object and was presented within a carrier phrase in sentence-medial position. Group performance was significantly above chance for three contrasts. However, the cochlear implant listeners performed much worse than a $6\frac{1}{2}$-year-old, moderately hearing impaired control listener who was tested on the same task. These findings suggest that this group of cochlear implant users had great difficulty in extracting the pitch information needed to accurately identify Cantonese lexical tones. [DOI: 10.1121/1.1471897]

PACS numbers: 43.71.Ky, 43.71.Hw, 43.66.Ts [CWT]

I. INTRODUCTION

Several investigations of pitch perception by cochlear implantees have studied the pitch percepts generated by stimulating the electrodes of multi-channel cochlear implants (see, e.g., Busby *et al.*, 1994; Busby and Clark, 2000; Collins *et al.*, 1997; Nelson *et al.*, 1995; Zwolan *et al.*, 1997). While individual differences among cochlear implantees have been reported in all studies, for a majority of subjects the pitch percepts changed from low to high as the position of the stimulated electrodes moved from the apex to the base of the cochlea in a manner similar to the tonotopic organization of pitch percepts in the normal ear.

For listeners with normal hearing, the pitch of complex sounds (called "pitch," hereafter) is determined mainly on the basis of the frequency of resolved, low-numbered harmonics (Moore *et al.*, 1985; Plomp, 1967; Ritsma, 1967). Dai (2000) demonstrated that harmonics in the vicinity of 600 Hz carry the largest weight in the calculation of pitch for normal-hearing subjects. While the frequencies of resolved harmonics are likely to be the most important cues for pitch perception, it is also possible to obtain a pitch percept, albeit an ambiguous one, from unresolved harmonics (Schouten *et al.*, 1962). Although the presence of a tonotopic organization of pitch percepts is likely to result in the accurate perception of the pitch of pure tones and narrow-band stimuli, it is not well understood how the pitch of complex stimuli is perceived through the electrical stimulation of the cochlea.

The most common processing strategy used by current cochlear implants is the "continuous interleaved sampling" (CIS) method (Wilson *et al.*, 1991). This strategy represents complex sounds as a set of amplitude-modulated signals presented through an array of electrodes that are placed within the cochlea. The signals presented through each electrode consist of a carrier pulse that has a 1 to 2 kHz frequency, and whose modulation rate typically preserves temporal information below 400 Hz. Only one electrode is stimulated at any time to prevent interaction of the electrical fields of adjacent electrodes. Faulkner *et al.* (2000) pointed out that listeners who use CIS cochlear implants are not able to resolve low-numbered harmonics of complex sounds whose fundamental frequencies are within the typical range of speech sounds due to the relatively wide bandpass filters used to deliver electrical stimulation to each electrode. Therefore, CIS users should not be able to perceive pitch on the basis of the frequencies of low-numbered, resolved harmonics of complex sounds. CIS implant users might only be able to make use of the weak cues provided by the periodicity information from unresolved harmonics and by overall differences in the amplitude of stimulation across different channels (Geurts and Wouters, 2001). Therefore, cochlear implantees are likely to have difficulties in perceiving the pitch of quasi-periodic sounds like speech and music (Faulkner *et al.*, 2000). Retrieval of within-channel periodicity information should be even more difficult for listeners using cochlear implants which employ low-pulse rate processing strategies such as the SPEAK processing strategy implemented on Nucleus 22 or 24 implants, which employs a pulse rate per channel that is typically lower than 250 Hz (McKay and McDermott, 1993).

While for nontonal languages the availability of auditory cues to pitch perception does not affect the performance on vowel and consonant recognition tasks (Faulkner *et al.*, 2000), different results might be obtained for the perception of languages in which pitch information is used in a contrastive way to cue lexical meaning (Lee and Nusbaum, 1993; Repp and Lin, 1990). For example, in Mandarin each syllable has one of four tones which differ primarily in $F0$ con-

[a)]Electronic mail: vciocca@hkusua.hku.hk

tour and level: tone 1 has a relatively high and flat *F*0 contour, tone 2 has a rising contour, tone 3 has a falling and rising contour, and tone 4 has a falling contour. Fu *et al.* (1998) showed that the perception of Mandarin tones can be advantageous for the accurate perception of segmental information. They asked native speakers of Mandarin with normal hearing to perform consonant, vowel, tone, and word recognition tasks by using processed speech. The speech signals were filtered through one, two, three, or four frequency bands. The signal within each band was then half-wave rectified and low-pass filtered at either 50 or 500 Hz in order to remove spectral information within each frequency band while preserving temporal envelope cues. This processing is similar to that of CIS processors, although the most recent versions of the latter employ a larger number of frequency bands (between 8 and 20 bands; see Loizou, 1998, for a review). The results showed that performance improved when the number of frequency bands was increased for the vowel, consonant, and word recognition tasks, but not for the tone recognition task. The advantage of tone recognition in segmental speech perception was shown by the fact that in the one-band, 500-Hz low-pass filtering condition, Mandarin listeners performed better (11% correct) than English listeners (2.9%; Shannon *et al.*, 1995) in a similar word recognition task. In other words, when spectral information is extremely limited, the ability to recognize tones gives an advantage in the recognition of vowels and consonants.

Another important finding of Fu *et al.*'s (1998) study concerns the acoustic cues that can be used for Mandarin tone recognition. They found that performance in tone recognition was well above chance for all conditions, but was affected by the low-pass filtering condition. As expected, the 500-Hz condition produced significantly better tone recognition than the 50-Hz condition. The finding that performance was well above chance even in the 50-Hz condition, for which the temporal envelope cues did not include periodicity information, suggests that listeners might have used temporal envelope cues such as stimulus duration and amplitude contour for identifying tones. This possibility is supported by the finding that both tone and word recognition were highest for tones 3 and 4, for which the *F*0 and the amplitude contours were highly correlated. Listeners were apparently able to recognize these tones with a great degree of accuracy purely on the basis of temporal envelope cues. These results are in agreement with previous findings that Mandarin tones can be recognized on the basis of cues other than *F*0 contour and height, although it is widely recognized that *F*0 contour and height are the main cues to Mandarin tone recognition (Tseng and Massaro, 1986; Whalen and Xu, 1992).

The Cantonese tonal system differs from that of Mandarin in a number of ways. First, Cantonese has six contrastive tones defined according to their pitch height and contour: high level (HL), high rising (HR), mid level (ML), low falling (LF), low rising (LR), and low level (LL). There is also a high falling (HF) tone which does not usually appear in the Cantonese spoken in Hong Kong (Bauer and Benedict, 1997). Second, Cantonese tones have been found to be cued almost exclusively by *F*0 contour and height (Fok Chan, 1974; Vance, 1976). Therefore, Cantonese tones are ideal stimuli for testing the capacity of cochlear implant listeners to estimate the fundamental frequency of phonation for the purpose of perceiving the pitch patterns of speech sounds.

The goal of this study was to investigate the identification of Cantonese tones by early-deafened listeners with cochlear implants. As Busby and Clark (2000) pointed out, the study of early-deafened cochlear implantees can give an insight on the effects of stimulus deprivation on the later development of perceptual skills. In particular, Busby and colleagues found that cochlear implant users who experienced auditory deprivation early in development performed worse on an electrode trajectory discrimination task than implantees who received auditory stimulation at an early age (Busby and Clark, 1996; Busby *et al.*, 1993). They argued that this finding may be related to the increased neural atrophy that results from the lack of stimulation in the developing auditory system. Although this hypothesis would suggest that early-deafened cochlear implantees should have difficulty in the perception of the pitch of complex sounds, the studies by Busby and colleagues used direct electrode stimulation rather than acoustic stimuli to test pitch perception. In the present study, the main goal was to determine whether early-deafened cochlear implantees could extract pitch information from natural speech sounds in order to recognize Cantonese tones.

II. METHOD

A. Subjects

Seventeen native Cantonese-speaking children (nine females, eight males) aged between 4 and 9 years old participated in this study. Only children older than the age of 4 years were included because Ching (1984) showed that even normal-hearing children are unable to reliably recognize isolated lexical tones until age 4. The reported onset of deafness ranged from birth to 30 months; the age at which the implants were fitted varied between 2 years and 6 months and 7 years and 7 months. Early-deafened children using cochlear implants have been found to require at least 12 months experience to have performance level of above 50% accuracy in English lexical stress recognition tasks and close-set word recognition tests (Tyler *et al.*, 1997). Therefore, all listeners had a postsurgical period longer than 12 months except for S1 (11 months), with a range of 11 to 41 months. All children used the Nucleus 24 cochlear implant system (™Cochlear Limited), except for one who used the Nucleus 22 implant. Six employed the SPEAK processing strategy (Seligman and McDermott, 1995), and 11 used Cochlear Limited's ACE speech processing strategy (see Kiefer *et al.*, 2001). For ACE users, the pulse rate was either 1200 Hz (four children), 900 Hz (six children), or 720 Hz (one child). The pulse rate was 250 Hz for all SPEAK users. Participants were fitted with cochlear implants either at the Queen Elizabeth, the Prince of Wales, or the Queen Mary Hospitals in Hong Kong, where they also received auditory and speech training.

B. Stimuli

Natural stimuli were used because children respond best to natural speech tokens in lexical tone identification tasks

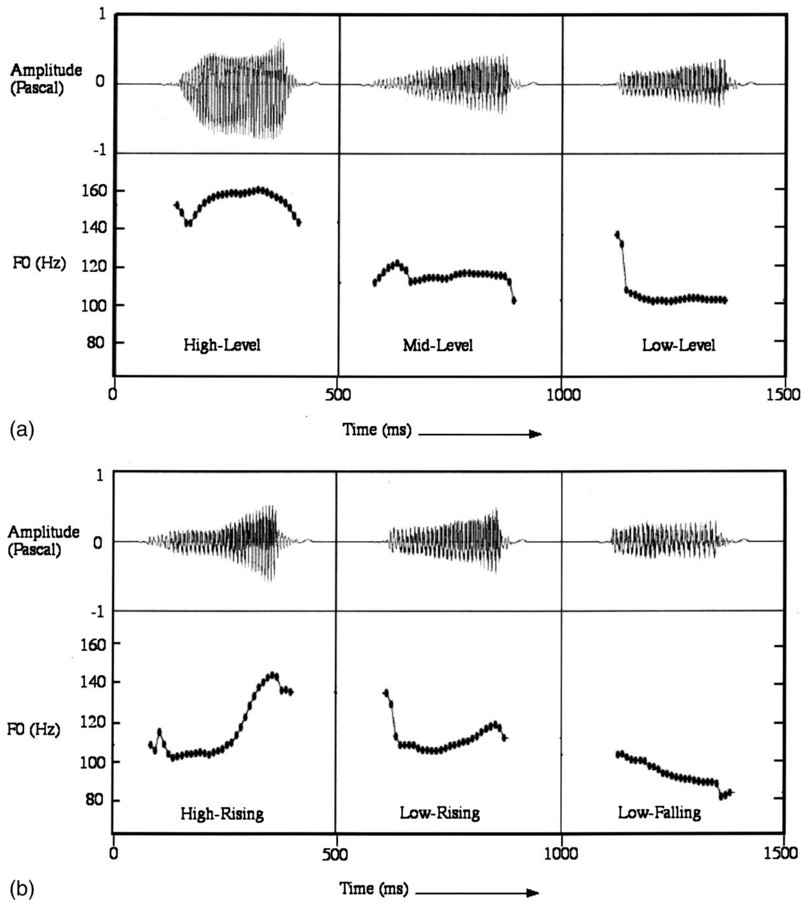

FIG. 1. Amplitude waveforms (top display) and corresponding *F*0 patterns (bottom displays) of the three level tones (a) and of the three contour tones (b) employed in the present study.

(Ching, 1984). The segmental sequence /ji/ was chosen as the basis for the target words as it can be represented by simple and concrete lexical items when produced with any of the six contrastive tones of Cantonese: high-level, /ji55/ (clothing); high-rising, /ji25/ (chair); mid-level, /ji33/ (spaghetti); low-falling, /ji21/ (child); low-rising, /ji23/ (ear); low-level, /ji22/ (two) (see footnote 1). All stimuli were produced by a native Cantonese male speaker aged 21, and recorded at a mouth-to-microphone distance of about 10 cm. The utterances were recorded onto the hard disk of an Apple PowerMacintosh 7100/AV using a Bruel & Kjaer Type 4003 microphone and a Type 2812 MK II microphone preamplifier. The six words were produced ten times each in random order within the carrier phrase:/η⊃23 wui23 t⊃k22 ___ pei25 lei23 teη55/ ("I will read ___ for you to hear"). The carrier phrase contained the target words in medial position to mitigate the influence of sentential intonation that might affect the fundamental frequency range of a word in initial or final position (Vance, 1976). The sentence with the smallest total difference from the average *F*0 calculated across all instances of each word was designated as the "context" sentence. The productions of each target word with the most extreme *F*0 difference, in the appropriate direction, from the mean *F*0 values of each word were used as the target stimuli. These stimuli were digitally clipped out of their respective sentence, normalized in amplitude, and digitally inserted into the context sentence. Informal listening tests were conducted using native Cantonese listeners with normal hearing to ensure that the carrier sentence with each of the target stimuli sounded natural in terms of prosody and tones.

An acoustic analysis of the fundamental frequency (*F*0) patterns of the target stimuli was conducted with the autocorrelation algorithm of the PRAAT software[2] in order to measure the range of *F*0 variation both within and across stimuli. The results of the *F*0 analyses for each level and rising/falling tones are shown in Figs. 1(a) and (b), respectively, together with the amplitude waveforms for each of the target stimuli. The *F*0 patterns and durations of the target stimuli are typical of Hong Kong Cantonese tones (see Bauer and Benedict, 1997). The mid-level, high-rising, low-rising, low-falling, and low-level had a starting *F*0 of 100 to 110 Hz.[3] Four stimuli had about the same duration; they were the high-level (280 ms), the high-rising (277 ms), the low-rising (283 ms), and the low-falling (269 ms) tones. The mid-level and the high-rising tones had longer duration (336 and 337 ms, respectively). The amplitude envelope is rising over time for the mid-level, low-level, high-rising, and low-rising tones, while it is relatively steady state with short-duration onset and offsets for the high-level and low-falling tones. It is important to stress that for normal-hearing listeners the duration and amplitude envelope cues are not important for the perception of tonal identity in Cantonese (Fok Chan, 1974).

The six target stimuli were grouped into the following eight tonal contrasts: (*i*) high-level versus mid-level ("HL-ML"; tone 55 vs 33), (*ii*) high-level versus low-level ("HL-

LL''; tone 55 vs 22), (*iii*) mid-level versus low-level (''ML-LL''; tone 33 vs 22), (*iv*) high-rising versus low-rising (''HR-LR''; Tone 25 vs 23), (*v*) low-rising versus low-level (''LR-LL''; tone 23 vs 22), (*vi*) low-falling versus low-rising (''LF-LR''; tone 21 vs 23), (*vii*) low-falling versus low-level (''LF-LL''; tone 21 vs 22), (*viii*) high-level versus high-rising (''HL-HR''; tone 55 vs 25). Contrasts HL-ML, HL-LL, and ML-LL were used to investigate the effect of separation between the three pitch levels (high, mid, and low) on tone perception. Pairs HR-LR, LR-LL, LF-LR, and LF-LL were used to test listeners' sensitivity to *F*0 differences in the endpoint of tones, since the tones for these pairs start at similar frequencies but end at different frequencies. Tones in pair HR-LR have the same (rising) contour while tones for pairs LR-LL, LF-LR, and LF-LL have different contours. Finally, pair HL-HR contains tones that have a similar *F*0 endpoint but different initial *F*0.

C. Procedure

Children were tested individually in a double-walled IAC soundproof room. One experimenter sat inside the soundproof room behind the listener, while the another experimenter and the care-giver sat outside the soundproof room. A computer (Power Macintosh 7100/80AV) placed outside the soundproof room and running a Hypercard 2.4 program was used to present the visual and auditory stimuli. Each trial began with the presentation of a target word within the carrier phrase. After this, two pictures of real objects were displayed side-by-side; the pictures represented the two members of a given tonal contrast. Pictures were matched in size (width and height), and in distance from the observer. Visual stimuli were projected from the computer onto a screen placed in the soundproof room using a CTX EzPro 500 projector; projected images were approximately 0.5 m by 0.5 m. The subjects were given the following instructions: ''You will hear each word once, then you should point to one of the two pictures to tell me which word you have heard.'' An experimenter sat outside the soundproof room and recorded the selected response. Participants were encouraged to guess if they were not sure about the correct response. Each contrast was presented four times within a block of trials, twice with each target word. For each target word within a contrast, one trial had the pictures in one order (one picture on the left and the other on the right), the other trial had them in the opposite order. Each participant completed four blocks of trials; each block consisted of 32 tonal contrasts (four trials for each of the eight tonal contrasts). The order of presentation of the stimuli was randomized for each block of trials. The auditory stimuli were output through an Audiomedia II D/A board into a Madsen 0B822 audiometer, and then through a Westra LAB-501 loudspeaker. Participants sat 1 m away from the speaker in the soundproof room.

Before the experimental session, all participants were given 10 to 15 practice trials to ensure that (*i*) they were familiar with all the lexical items and the corresponding pictures, and (*ii*) they understood the nature of the task. These trials were identical to the experimental trials except that the experimenter explained each step of the task, gave feedback, and answered questions. No feedback was provided during the experimental session. Participants were allowed to take a short break whenever they requested it.

The level of the stimuli was measured as peak dBA level at the listening position with a sound-level meter (Bruel & Kjaer, Type 1625). Although each target stimulus was normalized, the peak dBA level for the six target stimuli varied within a range of 8 dBA. The listening level was therefore set such that the range was centered around 65 dBA. At this setting the levels of the target stimuli were the following: 69 dBA for tone 55, 66 dBA for tone 25, 60 dBA for tone 33, 63 dBA for tone 21, 63 dBA for tone 25, and 61 dBA for tone 22. The difference in dBA level among the target stimuli could in principle be a confounding factor in the experiment. However, an informal listening test showed that the perceived loudness of the stimuli was less variable than the dBA readings might suggest. Therefore, it is unlikely that a difference in amplitude level among the stimuli could have been used as a strategy to improve the performance of this task (the results of the experiment fully support this statement).

Before the testing of the cochlear implantees, a moderately hearing impaired child (aged 6 years and 6 months) was used as a control listener. He was wearing a hearing aid and completed a pilot test of the experimental procedure to determine whether the task could be accomplished by a child with hearing impairment (pure tone threshold average) comparable to that of individuals fitted with cochlear implants (Ciocca *et al.*, 2000). The results of this pilot study showed that the moderately hearing impaired listener performed at 92% correct or above on all contrasts except for the HR-LR contrast (58% correct), as expected for a child with normal hearing of the same age (Lui, 2000).

This study was carried out in conjunction with another study (Wong, 2000) that involved tone discrimination and tone identification tasks using other stimuli. Subjects were tested in the order of (a) I1-D-I2 or (b) I2-D-I1, where I1 was the tone identification task of this study while D and I2 were the tone discrimination task and the tone identification task of the other study. Nine subjects were tested in order (a) and eight subjects were tested in order (b).

III. RESULTS

The data were analyzed by computing the percentage of correct scores for each tonal contrast and for each subject. Response rates for individual tones were not calculated because these scores were highly dependent on the choice available to the listeners for a given contrast. For example, the performance for the mid-level tone is likely to be better when it is contrasted with the high-level tone (large *F*0 separation) than when it is paired with the low-level tone (small *F*0 difference).

The average correct scores for each tone contrast ranged from 61% to 50% (see Fig. 2). As a group, the children performed above chance for contrasts HL-ML, HL-LL, and HL-HR (binomial test; $N=272$, $p=\frac{1}{2}$, $\alpha=0.05$). However, even for these contrasts, only a few of the children performed above chance (75% or better) by a binomial test ($N=16$, $p=\frac{1}{2}$, $\alpha=0.05$). Four listeners (S7, S11, S12, and S17) performed above chance for the HL-ML contrast. Only two

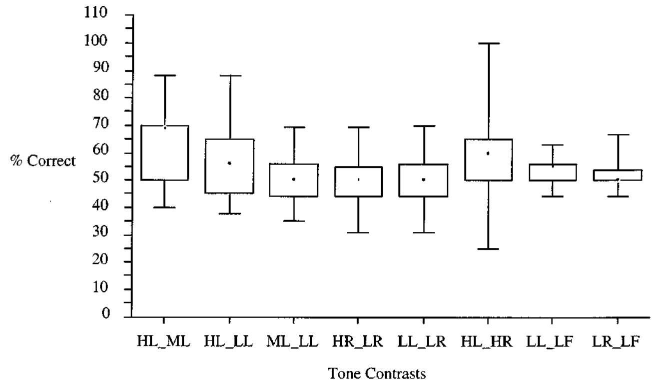

FIG. 2. Box plot of the performance (% correct) of the cochlear implant listeners, showing the median, minimum, and maximum values, and the first and third quartiles for each tonal contrast.

listeners (S12 and S15) performed above chance for contrasts HL-LL and HL-HR. None of the children performed above chance for any of the other five contrasts. Indeed, only 2 out of 17 listeners performed above chance overall (binomial test; $N=128$, $p=\frac{1}{2}$, $\alpha=0.05$).

A one-way ANOVA with repeated measures was carried out on the mean percent correct identification for each subject and each contrast. The results of the ANOVA showed that the means of tonal contrasts were significantly different, $F(7,112)=2.71$, $p<0.05$. Specifically, the performance for contrast HL-ML was significantly better than that for contrasts ML-LL and HR-LR (*post-hoc* Tukey HSD tests, $p<0.05$). None of the other pairwise comparisons between contrasts were statistically significant (Tukey HSD tests, $p>0.05$). These findings suggest that listeners tend to be more accurate at recognizing tones when the alternative choices differed by a large *F*0 separation (HL-ML contrast) than when the *F*0 separation was small (ML-LL) or the *F*0 contours were very similar (HR-LR). An exception to this hypothesis could be the finding that the HL-LL contrast was not perceived with significantly higher accuracy than the ML-LL contrast. However, it is important to notice that performance for the HL-LL, but not the ML-LL, was significantly better than chance.

Given that only two listeners performed above chance overall, it is perhaps not surprising that the correlations between overall performance and age at testing ($r=0.05$, $p>0.05$), duration of the postsurgical period ($r=-0.01$, $p>0.05$), age at implantation ($r=0.1$, $p>0.05$), and onset of deafness ($r=0.1$, $p>0.05$) were not statistically significant. Furthermore, the two listeners who performed best overall (S12 and S15) did not exhibit extreme values that might suggest any trend for any of the above variables.

IV. GENERAL DISCUSSION

Individual results for the tonal contrasts show that very few cochlear implant listeners performed above chance in a tone identification task in which they had to choose between two minimal pair alternatives. As a group, performance was above chance for three out of eight contrasts (HL-ML, HL-LL, and HL-HR), but did not exceed 61% correct on any contrast. These results suggest that early-deafened cochlear implantees have great difficulty in extracting *F*0 information on the basis of the input provided by cochlear implants. The results of a tone discrimination task performed by Wong (2000) using the same listeners further support this suggestion. She presented 30 same/different tone pairs produced by a male speaker in isolation with the syllable/wai/, and found that the overall group performance was 59% correct. Although this performance is above chance by a binomial test ($N=510$, $p=\frac{1}{2}$, $\alpha=0.05$), it does not represent a very accurate performance. Moreover, only four listeners performed significantly above chance in this task by a binomial test ($N=30$, $p=\frac{1}{2}$, $\alpha=0.05$).

Interestingly, all three contrasts that were identified above chance had the high-level tone as one of the members of the pairs. It is possible that group performance in contrast with the high-level tone was better because of the relatively large *F*0 separation between this tone and the other tones. For example, the average *F*0 separation in the level portion of the tones was about 45 Hz between the high-level and the low-level tones, and about 35 Hz between the high-level and the mid-level tones. These separations are well above the *F*0 difference threshold for fundamental frequencies around 150 Hz for CIS implant users (Geurts and Wouters, 2001). On the other hand, the contrast between mid-level and low-level tones was not perceived above chance. These tones were separated over most of their duration by an *F*0 difference (about 10 Hz) which is close to the *F*0 difference threshold for these listeners (Geurts and Wouters, 2001). An alternative reason for the better performance on contrasts involving the high-level tone could be that for some, but not all, speakers this tone has been found to have a higher overall amplitude level than the other tones (Fok Chan, 1974, pp. 139–148). Although this feature of the high-level tone is not produced consistently by all Cantonese speakers, some of the early-deafened children might have learned to exploit this potential cue for the identification of the high-level tone, and they might have used overall amplitude level as a cue for identifying the high-level tone in this experiment. This possibility is supported by the fact that the contrast between low-rising and low-falling tones was not identified above chance. For this contrast, the *F*0 difference at the offset is relatively large (about 40 Hz) but the overall amplitude level of the two tokens is similar. Other potential cues to the identification of tones could be the shape of amplitude envelope and the overall duration. However, Fok Chan (1974) did not identify any amplitude or duration pattern that was consistently associated with specific tones. Although in the current stimuli there were differences in overall amplitude, amplitude envelope, and duration, such differences are not consistently associated with lexical tone differences in the ambient language. Therefore, it is unlikely that the children in this experiment learned to use cues unrelated to pitch to identify Cantonese lexical tones, even though their use might have proven effective with the stimuli employed in this study.

The relatively poor performance of early-deafened cochlear implant listeners could be accounted for by several factors, including the etiology of deafness, the age of the child, the age of implant fitting, and the duration of the post-operative period. However, none of these variables was found to correlate with identification performance. These results are in apparent contrast with the claim by Busby and Clark (2000) that the duration of auditory deprivation prior to implantation is inversely correlated to performance on an electrode trajectory discrimination task. They found that the implantees in their worst performing group (S11, S15, S17, and S18) also had significantly longer duration of auditory deprivation, and older age at implantation and at testing than the other two groups. However, these group differences do not entirely account for differences in individual performance. For example, their subjects S6, S13, S14, and S16, who belonged to the highest performing group, had a duration of auditory deprivation which was longer (9 years and 5 months or more) than that of their subject S11 (7 years and 6 months). On the other hand, it should also be pointed out that the range of auditory deprivation of the listeners in the present study (1 year and 4 months to 7 years and 1 month) was considerably smaller than that of Busby and Clark's (1 year and 3 months to 18 years and 6 months). Therefore, it is possible that any effects of duration of auditory deprivation would be more difficult to observe in the current study.

The present findings are also in apparent contrast with studies on Mandarin tone perception by cochlear implantees. Huang *et al.* (1995) asked Mandarin-speaking adult implantees who were fitted with the Nucleus 22 implant to perform a four-alternative, forced choice tone recognition task. Their listeners were able to perceive Mandarin tones with about 68% accuracy, compared to a preoperative performance of 34.5% correct. Huang and colleagues suggested that the acoustic cues to fundamental frequency of the four Mandarin tones can be extracted by the speech-coding strategy of the Nucleus 22 cochlear implant system and stimulate the auditory nerve where they are perceived as pitch. There are two differences between the two studies that could account for the seemingly contrasting findings between Huang *et al.*'s and the present study. First, the Mandarin implantees were adults who had auditory and phonological knowledge before the implantation, and therefore might have been able to benefit more from the auditory capacity provided by the implant than early-deafened children (Boothroyd *et al.*, 1991). Adults may also perform better than children in a tone identification task because of more advanced cognitive skills in comparison with the young children. Second, mandarin speakers can make use of temporal envelope cues to recognize lexical tones (Fu *et al.*, 1998; Tseng and Massaro, 1986; Whalen and Xu, 1992). Therefore, the better performance of Mandarin cochlear implantees may have to do with the perception of temporal cues rather then pitch perception. By contrast, it has been shown that such nonpitch cues are unlikely to be reliable cues to tone perception in Cantonese (Fok Chan, 1974; Vance, 1976).

Given that Cantonese cochlear implant listeners are not likely to use temporal envelope cues for tone recognition, their recognition of lexical tones is most probably based on pitch perception. However, the perception of pitch through current multi-channel cochlear implants cannot be accomplished through information about the frequencies of low-numbered, resolved harmonics. Instead, cochlear implantees may rely upon periodicity cues resulting from the interaction of two or more unresolved harmonics. Since periodicity information is generally considered to be a weak cue to pitch, it is perhaps not surprising that early-deafened Cantonese cochlear implant users have difficulties in recognizing lexical tones. Given that even the listeners with higher pulse rates (900 or 1200 Hz) had extreme difficulties in recognizing the lexical tones, it is unlikely than the failure to use periodicity cues for pitch perception was due to a lack of periodicity information transmitted by the processors of the cochlear implants.

Although the performance of this group of cochlear implant listeners was poor overall, it is not possible to know whether these children will be able to improve their Cantonese tone perception skills in the future or whether the relatively impoverished auditory input they receive through the implants will not allow them to learn to identify Cantonese tones in a consistent way. Further studies will have to be carried out in order to determine whether the quality of the auditory input or other cognitive and/or linguistic factors are likely to be the main contributors to the lexical tone perception abilities of Cantonese-speaking cochlear implant listeners.

ACKNOWLEDGMENTS

The authors wish to thank Dr. H.C. Yu and Eddie Wong, ENT Department, Queen Elizabeth Hospital, Hong Kong; Terence Wong, Department of Surgery, Prince of Wales Hospital, Hong Kong; Bessie Pang, Suen Mei Speech and Hearing Center for the Deaf, Hong Kong; Raymond Wu and Donald Chan for technical support; and all participants and their families for their cooperation.

[1]The two digits following each word represent the starting and ending points of the pitch contour of each tone, according to the system presented by Bauer and Benedict (1997), where "1" represents the lowest pitch and "5" is the highest pitch of a talker's conversational pitch rang. For example, the high-rising tone represented as 25 has a starting point at the lower end of a talker's pitch range (2), and terminates at the highest end of the range (5).

[2]PRAAT 3.9.27 Copyright © 1992–2001 by Paul Boersma and David Weenink (www.praat.org).

[3]The first two pitch pulses of the low-level and low-rising tones had an $F0$ of about 130 Hz. Relatively large variations in $F0$ at the onset and offset of the syllables of tonal languages are not uncommon (see Bauer and Benedict, 1997). However, these short-term variations in $F0$ are typically not perceived as deviations from the overall pitch pattern of the tones. In this case, the low-level and low-rising tones were perceived as having a relatively low starting pitch by native Cantonese listeners with normal hearing.

Bauer, R. S., and Benedict, P. K. (**1997**). *Modern Cantonese Phonology* (Mouton de Gruyter, Berlin).

Boothroyd, A., Geers, A. E., and Moog, J. S. (**1991**). "Practical implications of cochlear implants in children," Ear Hear. **12**, 81–89.

Busby, P. A., and Clark, G. M. (**1996**). "Electrode discrimination by early-deafened cochlear implant patients," Audiology **35**, 8–22.

Busby, P. A., and Clark, G. M. (**2000**). "Pitch estimation by early-deafened subjects using a multiple-electrode cochlear implant," J. Acoust. Soc. Am. **107**, 547–558.

Busby, P. A., Tong, Y.-C., and Clark, G. M. (**1993**). "Electrode position, repetition rate, and speech perception by early- and late-deafened cochlear implant patients," J. Acoust. Soc. Am. **93**, 1058–1067.

Busby, P. A., Whitford, L. A., Blamey, P. J., Richardson, L. M., and Clark, G. M. (**1994**). "Pitch perception for different modes of stimulation using the Cochlear multiple-electrode prosthesis," J. Acoust. Soc. Am. **95**, 2658–2669.

Ching, T. Y. C. (**1984**). "Lexical tone pattern learning in Cantonese children," Lang. Learn. Comm. **3**(3), 317–334.

Ciocca, V., Aisha, R., Francis, A., and Wong, L. (**2000**). "Can Cantonese children with cochlear implants perceive lexical tones?" Paper presented at the International Conference on Speech and Language Processing, ICSLP2000, Beijing.

Collins, L. M., Zwolan, T. A., and Wakefield, G. H. (**1997**). "Comparison of electrode discrimination, pitch ranking, and pitch scaling data in postlingually deafened adult cochlear implant subjects," J. Acoust. Soc. Am. **101**, 440–455.

Dai, H. (**2000**). "On the relative influence of individual harmonics on pitch judgment," J. Acoust. Soc. Am. **107**, 953–959.

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.

Fok Chan, Y. Y. (**1974**). *A Perceptual Study of Tones in Cantonese* (Centre of Asian Studies, University of Hong Kong, Hong Kong), Vol. 18.

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.

Huang, T.-S., Wang, N.-M., and Liu, S.-Y. (**1995**). "Tone perception of Mandarin-speaking postlingually deaf implantees using the nucleus 22-channel cochlear mini system," Ann. Otol. Rhinol. Laryngol. Suppl. **166**, 294–298.

Kiefer, J., Holl, S., Stürzebecher, E., Pfennigdorff, T., and Gstöettner, W. (**2001**). "Comparison of speech recognition with different speech coding strategies (SPEAK, CIS, ACE) and their relationship to telemetric measures of compound action potential in the nucleus CI 24M cochlear implant system," Audiology **40**, 32–42.

Lee, L., and Nusbaum, H. C. (**1993**). "Processing interactions between segmental and suprasegmental information in native speakers of English and Mandarin Chinese," Percept. Psychophys. **53**, 157–165.

Loizou, P. C. (**1998**). "Introduction to cochlear implants," IEEE Signal Process. Mag. **15, 5**, 101–130.

Lui, J. (**2000**). "Cantonese tone perception in children," unpublished B.Sc. dissertation, University of Hong Kong.

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.

Moore, B. C. J., Glasberg, B. R., and Peters, R. W. (**1985**). "Relative dominance of individual partials in determining the pitch of complex tones," J. Acoust. Soc. Am. **77**, 1853–60.

Nelson, D. A., Van Tasell, D. J., Schroder, A. C., Soli, S., and Levine, S. (**1995**). "Electrode ranking of 'place pitch' and speech recognition in electrical hearing," J. Acoust. Soc. Am. **98**, 1987–1999.

Plomp, R. (**1967**). "Pitch of complex tones," J. Acoust. Soc. Am. **41**, 1526–1533.

Repp, B. H., and Lin, H.-B. (**1990**). "Integration of segmental and tonal information in speech perception: A cross-linguistic study," J. Phonetics **18**, 481–495.

Ritsma, R. J. (**1967**). "Frequencies dominant in the perception of the pitch of complex sounds," J. Acoust. Soc. Am. **42**, 191–198.

Schouten, J. F., Ritsma, R. J., and Cardozo, B. L. (**1962**). "Pitch of the residue," J. Acoust. Soc. Am. **34**, 1418–1424.

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

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

Tseng, C.-Y., and Massaro, D. W. (**1986**). "Lexical tone perception in Mandarin Chinese: Evaluation and integration of acoustic features," in *Linguistics, Psychology, and the Chinese Language*, edited by A. K. a. R. Hoosain (Hong Kong U.P., Hong Kong), pp. 91–104.

Tyler, R., Parkinson, A. J., Bertchy, H. F., Lowder, M. W., Parkinson, W. S., Gantz, B. G., and Kelsay, D. M. (**1997**). "Speech perception by prelingually deaf children and postlingually deaf adults with cochlear implant," Scand. Audiol. Suppl. **26**, (Suppl 46), 65–71.

Vance, T. J. (**1976**). "An experimental investigation of tone and intonation in Cantonese," Phonetica **33**, 368–392.

Whalen, D. H., and Xu, Y. (**1992**). "Information for Mandarin tones in the amplitude contour and in brief segments," Phonetica **49**, 25–47.

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 implant," Nature (London) **352**, 236–238.

Wong, A. (**2000**). "Tone perception performance of prelingually hearing-impaired Cantonese-speaking children with cochlear implants," unpublished M.Sc. dissertation, University of Hong Kong.

Zwolan, T. A., Collins, L. M., and Wakefield, G. H. (**1997**). "Electrode discrimination and speech recognition in postlingually deafened adult cochlear implant subjects," J. Acoust. Soc. Am. **102**, 3673–3685.