Published in final edited form as:

Int J Audiol. 2014 August ; 53(8): 546–557. doi:10.3109/14992027.2014.893374.

The Effect of Enhancing Temporal Periodicity Cues on Cantonese Tone Recognition by Cochlear Implantees

Tan LEE^{a),*}, Shing YU^{a)}, Meng YUAN^{b),c)}, Terence Ka Cheong WONG^{d)}, and Ying-Yee KONG^{e)}

^{a)}Department of Electronic Engineering, The Chinese University of Hong Kong

^{b)}Shanghai Acoustics Laboratory, Institute of Acoustics, Chinese Academy of Sciences, Shanghai, 200032, China

^{c)}Key Laboratory of Speech and Hearing Sciences (East China Normal University), Ministry of Education, 200062 Shanghai, China

^{d)}Department of Otorhinolaryngology, Head and Neck Surgery, Institute of Human Communicative Research, The Chinese University of Hong Kong, Hong Kong

^{e)}Department of Speech Language Pathology & Audiology, Northeastern University, Boston, MA 02115, USA

Abstract

Objectives—This study investigates the efficacy of a cochlear implant (CI) processing method that enhances temporal periodicity cues of speech.

Design—Subjects participated in word and tone identification tasks. Two processing conditions – the conventional Advanced Combination Encoder (ACE) and tone-enhanced ACE were tested. Test materials were Cantonese disyllabic words recorded from one male and one female speaker. Speech-shaped noise was added to clean speech. The fundamental frequency information for periodicity enhancement was extracted from the clean speech. Electrical stimuli generated from the noisy speech with and without periodicity enhancement were presented via direct stimulation using a Laura 34 research processor. Subjects were asked to identify the presented word.

Study sample—Seven post-lingually deafened native Cantonese-speaking CI users.

Results—Percent correct word, segmental structure, and tone identification scores were calculated. While word and segmental structure identification accuracy remained similar between the two processing conditions, tone identification in noise was better with tone-enhanced ACE than with conventional ACE. Significant improvement on tone perception was found only for the female voice.

Conclusions—Temporal periodicity cues are important to tone perception in noise. Pitch and tone perception by CI users could be improved when listeners received enhanced temporal periodicity cues.

^{*)}Corresponding author: Address: RM324, Ho Sin Hang Engineering Building, The Chinese University of Hong Kong, Hong Kong, Tel: 852-39438267, Fax: 852-26035558, tanlee@ee.cuhk.edu.hk.

Keywords

Cochlear implant; Cantonese; pitch; tone perception; periodicity enhancement

Introduction

Cochlear implants (CI) are electronic systems designed to partially restore the auditory functions of individuals with severe-to-profound hearing impairment. In the past decade, the number of CI users worldwide has increased rapidly (NIDCD, 2009). In China, the survey of the National Bureau of Statistics of China estimated about 20 million confirmed cases of hearing disabilities, and many of these patients are potential CI recipients (China Disabled Persons' Federation, 2008). Although current CI devices are generally effective in giving users the ability to understand speech in quiet, many challenges remain in real-world speech communication. In particular, many CI users find it difficult to perceive pitch (e.g., Zeng, 2002; Kong et al., 2009), which is an essential cue for the perception of speech intonation (e.g., Chatterjee & Peng, 2008; Peng et al., 2008), understanding speech in competing backgrounds (e.g., Stickney et al., 2004; Fu & Nogaki, 2005), and the perception of musical melodies (e.g., Gfeller et al., 2002; Kong et al., 2004; Nimmons et al., 2008). Reduced pitch perception ability also has significant impacts on speech recognition for CI users who use tonal languages in daily communication (e.g., Lee et al., 2002; Ciocca et al., 2002; Wei et al., 2004; Xu and Zhou, 2012).

Pitch is a subjective attribute that is closely related to the fundamental frequency (F0) of voiced speech or tonal music signals. For normal-hearing listeners, both temporal and spectral cues contribute to complex pitch perception. In the spectral domain, pitch is encoded by resolving the low-order harmonics along the basilar membrane (spectral fine-structure cues) (e.g., Plomp, 1967; Ritsma, 1967; Houtsma and Smurzynski, 1990). In the temporal domain, F0 information is encoded via phase locking to individual harmonics (temporal fine-structure cues) (e.g., Schouten, 1940; Licklider, 1951; 1959; Meddis and O'Mard, 1997) or to the temporal periodicity of the amplitude-modulated envelopes of unresolved harmonics (temporal envelope-periodicity cues) (Burns and Viemesiter, 1976, 1981). It has been shown that pitch elicited by temporal and spectral fine-structure cues is more salient than that elicited by the envelope-periodicity cue (Smith et al., 2002).

However, most existing CI systems have very limited spectral resolution, making speech harmonics unresolvable (Qin & Oxenham, 2005). This is due to a combination of factors including current spread (e.g., Nelson et al., 2011), reduced neural survival, and the small number of electrodes. (e.g., Li & Fu, 2010). As for temporal coding of pitch, most clinical CI strategies discard the temporal fine-structure cue, which is the phase information (i.e., the carrier wave) of the signal, due to the use of a fixed-rate carrier. A typical method of decomposing a signal into temporal envelope and fine structure is the Hilbert transform (Hilbert, 1912). In the most commonly-used CI speech processing strategies, namely Continuous Interleaved Sampling (CIS) (Wilson et al., 1991) and Advanced Combination Encoder (ACE) (Vandali et al., 2000), the input signal is divided into a pre-defined number of frequency bands, from each of which the information related to the amplitude fluctuation

of the signal is extracted. Only the amplitude fluctuations that reflect the slowly-varying temporal envelope (2 – 50 Hz) and the pitch-related periodicity (50 – 500 Hz) are extracted (Rosen, 1992). Thus, the electrical stimuli transmitted to the CI users contain primarily temporal envelope cues for pitch and music perception, and for tone recognition in tonal languages. There have been many studies concerning the roles of temporal envelope and periodicity cues (TEPC) in pitch perception and lexical tone recognition. The contributions of TEPC to tone recognition of Mandarin (Kong and Zeng, 2006) and Cantonese (Yuan et al., 2009) were investigated by acoustic simulation with noise-excited vocoder speech. Both studies showed that, with only TEPC, tone recognition could reach a fairly good performance level in quiet with four bands of frequency information. However, TEPC was found to be highly susceptible to noise. Tone recognition accuracy with TEPC decreased significantly when the signal-to-noise ratio (SNR) decreased.

Different approaches have been suggested for improving the delivery of pitch information in CI systems. One approach is to refine the place pitch cue by designing band-pass filters with better spectral resolution in the low-frequency region such that the first harmonic is more likely to be resolved (Geurts and Wouters, 2004). This method was tested with a small number of CI users and improvement on F0 discrimination ability was observed. Another approach is to enhance temporal periodicity cues in the widely used CIS strategy and ACE strategy. Previous work by other researchers, such as Green and colleagues (2004; 2005), Wouters and colleagues (Geurts and Wouters, 2001; Laneau et al., 2006; Milczynski et al., 2009; Milczynski et al., 2012), and Vandali and colleagues (Vandali et al., 2005; Wong et al., 2008; Vandali and van Hoesel, 2011) has attempted to strengthen F0-related temporal envelope-periodicity cues to improve pitch and tone perception. These studies revealed two promising directions along which the design of CI processing strategies could be improved for better pitch perception. First, the sub-band temporal periodicity cues at individual electrodes could be made more salient, e.g., by increasing the modulation depth. Second, the F0-related periodic fluctuations should be synchronized across electrodes. A brief description of their work is provided below:

Green et al. (2004) proposed to use an F0-controlled sawtooth-like wave to replace the original periodicity component extracted from the input signal. The use of a simple periodic waveform eliminates the temporal complexities and ensures that speech periodicity is clearly represented. This method was implemented as a modified CIS strategy and tested with CI subjects. The F0-controlled strategy showed better performance in pitch glide labeling with synthetic vowels than the conventional CIS strategy. Green et al. (2005) further evaluated the F0-controlled strategy with natural speech stimuli. While its benefit to intonation perception was observed, vowel recognition was negatively affected.

Geurt and Wouters (2001) first investigated the effect of amplitude modulation depth on F0 discrimination and accordingly proposed a modified CIS algorithm with increased modulation depth. However, the modified algorithm did not show any benefit to CI users. Laneau et al. (2006) then proposed a new strategy in which the F0-related periodicity was represented by a sinusoidal wave. The processing strategy, named F0mod, was implemented by modifying the conventional ACE strategy. Test results with CI subjects showed the benefit of explicit pitch encoding in music perception. The F0mod strategy was further

developed by Milczynski et al. (2009) and tested on CI users. Significant improvement on pitch ranking and musical melody identification performance was demonstrated with the F0mod strategy compared with the conventional ACE strategy. Recently, Milczynski et al. (2012) examined the effectiveness of the F0mod strategy for Mandarin speech perception. Their results showed that F0mod performed better than conventional ACE for lexical tone recognition. For word and sentence recognition performance, they found no significant difference between the two strategies.

Vandali et al. (2005) developed the F0 Synchronized ACE (F0Sync) and Multi-channel Envelope Modulation (MEM) strategies, which aimed at enhancing F0 periodicity cues. In the F0Sync strategy, the stimuli on all active electrodes were synchronized using an F0-controlled gating signal. In the MEM strategy, synchronized F0 control across electrodes was realized by modulating the sub-band stimuli with the temporal envelope of the full-band input signal. They reported that both F0Sync and MEM provided significant benefits to CI users in pitch ranking tests compared to the ACE strategy. Wong et al. (2008) evaluated the effect of periodicity enhancement by the MEM strategy on Cantonese sentence recognition by CI users. The results did not show the benefit of MEM compared to ACE and CIS strategies. Recently, Vandali and van Hoesel (2011) presented the design of a new strategy, named eTone, which is a modified ACE with explicitly encoded F0 information. This strategy adjusted the contribution of the F0-modulation component in the output stimuli, based on the estimated degree of harmonicity at each individual channel. The eTone strategy was shown to produce a better representation of F0 timing information than ACE. Vandali and van Hoesel (2012) showed that the CI users performed better in a sung-vowel pitch-ranking task with the eTone strategy than with the ACE strategy.

Given the importance of envelope periodicity cues for music and pitch perception, we investigated the effect of explicit enhancement of temporal periodicity on Cantonese tone and speech perception by CI users. In our earlier work (Yuan et al., 2009), we reported perceptual data from normal-hearing listeners with CI simulation on Cantonese tone and word recognition in noise with and without periodicity enhancement. A four-channel noise-excited vocoder was used for simulating CI processing. In our periodicity enhancement method, the temporal periodicity cues in the range of 20 – 500 Hz at all channels were replaced by a sinusoidal wave, and the temporal envelopes below 20 Hz were unmodified. The frequency of the sinusoidal wave followed the F0 trajectory extracted from noise-free input signal prior to the test. The present study is an extension of Yuan et al. (2009), from acoustic simulation to tests with CI users. The simulation results showed consistent improvement of tone identification accuracy with the enhanced periodicity cues. The improvement was more significant at lower SNRs. Given that the ACE processing strategy is used by most of the available test subjects, we implemented the periodicity enhancement method by modifying the conventional ACE, making it similar to the F0mod strategy as described in Milczynski et al. (2012). Different from Milczynski et al. (2012), the present study tests perception of Cantonese tones instead of Mandarin tones. Cantonese is one of the major Chinese dialects, which is well known for tonal richness. Similar to Mandarin, in Cantonese, each character is pronounced as a monosyllable with a specific lexical tone. If the tone is different, the syllable refers to another character that has a different meaning. However, significantly different from Mandarin, which has only four tones, there are six

basic tones in Cantonese, which are characterized by distinctive pitch contours as illustrated in Figure 1. There are three level tones (tones 1, 3, and 6), two rising tones (tones 2 and 5), and one falling tone (tone 4) in this language. As an example, the syllable /fu/¹, when carrying different tones, can give six different meanings: “husband” (Tone 1), “tiger” (Tone 2), “wealthy” (Tone 3), “symbol” (Tone 4), “woman” (Tone 5), and “father” (Tone 6). Unlike Mandarin, the contribution of slow-varying temporal envelope for tone perception is very limited in Cantonese. For Mandarin, the temporal envelopes are distinctively different among four tones (see top panel of Figure 2). For example, tone 3 has a lower energy in the mid-portion of the syllable and tone 4 is generally shorter than the other tones. These differences allow the listeners to achieve high levels of accuracy for tone perception using only temporal envelope cues (e.g., Xu and Pfingst, 2003; Kong and Zeng, 2006). However, the temporal envelopes are similar across the six Cantonese tones (see bottom panel of Figure 2). Without temporal envelope cues, perception of the small differences in F0 for the three level tones would be particularly challenging for CI users. In other words, recognition of Cantonese level tones relies greatly on pitch cues, whereas the role of pitch perception is less critical in Mandarin tone identification. Thus, the current study on Cantonese tone perception would better reflect the effectiveness of periodicity enhancement to pitch perception than studies on Mandarin tones. Unlike Milczynski et al. (2012), who investigated mainly word and sentence recognition, the current study adopted the test paradigm proposed by Yuan et al. (2009) that independently evaluated tone, segmental structure, and word identification in the same test (see descriptions in the Methods section). Here the term “segmental structure” refers to the phonemic constituents that compose a word. For example, the Chinese character “父” (meaning “husband”) has a segmental structure of /fu/ and high-level tone (Tone 1). Using this paradigm, not only could the impact of periodicity enhancement on tone identification performance be evaluated directly, but also the impact on segmental structure recognition. In Milczynski et al. (2012), the dependence of word recognition performance on vowel and tone recognition was investigated using sensitivity analysis (Chen and Massaro, 2008).

In the following sections, we first describe the details of the signal processing and the listening test methods. We then present perceptual data obtained from a group of Cantonese-speaking CI users. The performance of the periodicity-enhanced method on Cantonese tone recognition and word recognition in noise are compared to that of the conventional ACE. The results showed significant periodicity enhancement benefit for Cantonese tone perception in noise.

Methods

Subjects

The listening tests in this study were carried out using the Laura 34 research processor provided by Cochlear Limited (Cochlear Limited, 2006). To be compatible with this experimental platform, the subjects must be implanted with the Nucleus CI systems from Cochlear Limited. Four male and three female subjects (S01–S07), aged 43 to 60 years

¹In this article, the Jyut Ping system is used for transcribing Cantonese syllables. Jyut Ping was devised by the Linguistic Society of Hong Kong (LSHK, 1997).

(mean = 49 years) participated in the listening tests. All of them were post-lingually deafened adults and native speakers of Cantonese with at least 6 months of experience in listening with a CI. All of the subjects use the ACE strategy in their daily communication. The brief demographic information of these subjects is provided in Table 1.

Test Materials

This study focuses on the performance of lexical tone identification by CI users. Lexical tones are carried by syllables that correspond to meaningful words in the language, and therefore the task of tone identification essentially becomes a task of word identification. In our previous study on acoustic simulation of a tone-enhanced processing method, Cantonese disyllabic words were used for tone identification by normal-hearing listeners (Yuan et al., 2009). In each test trial, the subject was presented with a disyllabic word. Four candidate disyllabic words were displayed on a computer screen and the subject was asked to identify the word he/she heard. To test tone perception, the candidate words were required to have the same segmental structure, but carry different tones. Given the lexical constraints of Cantonese, it was difficult to compose word sets that cover all of the six tones with the same syllable. Therefore, we chose to include only two contrasting tones in each test set for a given trial. Figure 3 shows one of the test set. In this example, the subject is presented with a disyllabic word /ging1 lik6/ (經歷). One of the candidate words /ging2 lik6/ (警力) shown on the screen contains a contrasting tone (i.e., tone 1 vs. tone 2 carried by the first syllables) with the same segmental composition. The other two disyllabic words, /gung1 lik6/ (功力) and /geng2 lik6/ (頸力), are used as distractors. They have the same intended tone contrast (i.e., tone 1 vs. tone 2), but slightly different segmental compositions (i.e., the syllables /gung/ and /geng/ are different from /ging/). The full set of stimuli used for this study can be found in Yuan et al. (2009) and Yuen et al. (2009). The speech materials were recorded from four native speakers (two male, two female) of Hong Kong Cantonese. The recording was carried out in a sound-treated booth. The recorded data were digitized at a sampling frequency of 44.1 kHz and a resolution of 16 bits. Recordings from one male (mean F0 = 115 Hz, range 81 – 188 Hz) and one female (mean F0 = 220 Hz, range 133 – 363 Hz) were used in the training session, whereas recordings from the remaining male (mean F0 = 104 Hz, range 76 – 158 Hz) and female (mean F0 = 183 Hz, range 129 – 254 Hz) were used for the testing session described below.

Noisy speech materials were created by adding a speaker-specific speech-spectrum-shaped noise to the clean speech. To generate the noise signal, all recorded speech materials from a particular speaker were divided into non-overlapping short-time frames, and the power spectrum of each frame was computed. The noise signal was produced by filtering a white noise to fit the average power spectrum over all speech frames. The noise was then added to each test utterance from the same speaker at different SNRs. With the root-mean-square (RMS) level of the target speech signal being fixed, the noise intensity was adjusted to achieve a prescribed SNR for the utterance. All stimuli were presented at the same RMS level; that is, no level roving was applied. Unlike pitch discrimination tasks reported in many psychophysical studies on CI, listeners in the current study were unlikely to use the loudness cues within each trial to perform tone identification. In our design of listening tests, the candidate words have different segmental structures and only one disyllabic word was

presented in each trial. Across trials, the perceived loudness of different disyllables could be very different, making the potential loudness cues across syllables unreliable for tone perception. Thus, level cues were unlikely to be an important factor affecting the test results. For each test stimulus, the noise was 1 second longer than the target speech, such that the first and the last 0.5 second of the signal contain only noise.

Signal Processing

The tone-enhanced ACE (henceforth abbreviated as ToneACE) was developed by modifying the ACE strategy, which is the clinical strategy used in the Nucleus CI systems by the seven test subjects. The implementation of conventional ACE is illustrated as in Figure 4 (Kiefer et al., 2001; Laneau, 2005). The input signal undergoes short-time spectral analysis with a frame rate of 500 to 1800 Hz, depending on the stimulation rate per channel for each subject. That is, the analysis frames advance with a time step of 0.6 to 2.0 ms. At each frame, a 128-point fast Fourier transform (FFT) is applied and the magnitudes of the resulting coefficients are combined to form 22 frequency channels, each corresponding to an implanted electrode. Typically 8 to 12 channels with the highest magnitude levels are used to generate electrical stimulation to the respective electrodes for this analysis frame, while the other electrodes do not deliver any stimulating signal. In each of the selected channels, the spectral magnitude is converted into an electrical stimulation level, which fits into the range between the Threshold Level (TL) and the Comfort Level (CL) of the subject. TL specifies the lowest stimulation level that elicits a soft but consistent auditory sensation. CL is defined as the maximum stimulation level above which uncomfortably loud sound would be produced (Cochlear Limited, 2009). Stimulations of all selected channels are presented to the respective electrodes in an interleaved manner from high-frequency channels to low-frequency channels.

Figure 5 illustrates the implementation of the ToneACE method. Similar to the conventional ACE, sub-band TEPC are extracted by short-time spectral analysis. Let $p_i[n]$ denote the TEPC of the i th channel, which is represented by 500 to 1800 samples per second. Low-pass filtering with 20 Hz cut-off frequency is applied to $p_i[n]$ such that the slowly varying temporal envelope, denoted by $e_i[n]$, is retained. The filter is a seventh-order elliptic filter with 0.3 dB of ripple in the pass-band, and stop-band attenuation of 80 dB. Subsequently, a periodicity-enhanced TEPC $p'_i[n]$ is obtained by using $e_i[n]$ to amplitude-modulate a sinusoidal wave as,

$$p'_i[n] = e_i[n] \times c[n],$$

where

$$c[n] = 1 + \sin \sum_{k=1}^n (2\pi F_0[k] \Delta t).$$

t is the sampling interval of the TEPC, and $F_0[k]$ is the F0 value in Hz of the input speech signal at the time instant corresponding to the k th sample of $e_i[n]$. $c[n]$ has a 100% modulation depth and is applied to all channels (i.e., synchronized across channels). In this way, the complex periodicity cue in the input speech is replaced by a simple periodic wave. It must be noted that, when processing noisy speech materials with ToneACE, we used the F0 trajectories pre-computed from the respective clean speech utterances. The pitch estimation algorithm in the PRAAT software (Boersma and Weenink, 2009) was used and the resultant F0 estimates were manually verified and corrected when necessary. Erroneous F0 values such as octave jump were re-estimated during this verification process.

Figure 6 shows an example to compare the original TEPC and the periodicity-enhanced TEPC extracted from a noisy speech signal. The waveforms of the original TEPC of the noisy signal ($p_i[n]$), the sinusoidal waveform that represents the F0 ($c[n]$), and the enhanced TEPC ($p'_i[n]$) are plotted in the upper, middle and lower panels, respectively. The slowly varying envelope $e_i[n]$ is superimposed as the thick solid curves on the plots of both $p_i[n]$ and $p'_i[n]$. All of these waveforms are obtained from a Cantonese word utterance with speech-spectrum-shaped noise at 0 dB SNR. Note that the low-pass filter introduces a time delay between $e_i[n]$ and $p_i[n]$. This delay was compensated for in the generation of $p'_i[n]$ by re-filtering in the reverse direction in time such that the entire processing has zero-phase distortion (Oppenheim et al, 1999; Geurt and Wouters, 2001). From the figure, it is noted that $p_i[n]$ does not exhibit clear periodicity because of the noise corruption. The periodicity is restored as the modulated sinusoidal wave in the enhanced TEPC. In this example, the F0 trajectory for generating $c[n]$ is estimated from clean speech. The TEPC for unvoiced and non-speech segments are kept unchanged. The effect of periodicity enhancement with ToneACE can also be illustrated by comparing the electrodograms generated from ACE and ToneACE (Figure 7). In this example, the Cantonese syllable /gei1/ spoken by a male speaker is corrupted by speech-shaped noise at 0 dB SNR. It can be seen that the electrodogram from ACE does not exhibit clear periodicity. With ToneACE, the temporal periodicity of the original speech in its voiced segment (0.55 – 0.95 sec) could be largely recovered.

As mentioned earlier, the design of ToneACE follows the same principle as the F0mod strategy in Laneau et al. (2006). The implementation of ToneACE in this study is similar to the modified version of F0mod as described in Milczynski et al. (2012), but there are some differences. First, for the extraction of slowly-varying temporal envelopes, an elliptic low-pass filter with cut-off frequency of 20 Hz was used in our implementation while a Butterworth low-pass filter with cut-off frequency of 60 Hz was used by Milczynski et al. Second, in ToneACE, the sinusoidal wave being modulated has an amplitude range of 0.0 to 2.0, and a DC level of 1.0. This is different from the implementation of Milczynski et al. (2012), where the amplitude range was 0.0 to 1.0 and the DC level was 0.5. This implies that the amplitude of TEPC of the F0-modulated signal in Milczynski et al. (2012) was smaller than that in this study. Third, for the generation of electrical stimuli, subject-specific processing parameters retrieved from the subjects' processors were used in our study, while the same processing parameters were applied to different subjects in Milczynski et al. (2012).

Setup

The test stimuli were generated by software-implemented signal processing algorithms. They were presented to a subject's implant via the Laura 34 research processor by using Nucleus Implant Communicator (NIC) software interface (Cochlear Limited, 2006). The Laura 34 research processor and the NIC software interface were developed and manufactured by Cochlear Limited. The software algorithms of stimuli generation were developed using the Nucleus MATLAB Toolbox (NMT) provided by Cochlear Limited (Swanson and Mauch, 2006). NMT includes the ACE algorithm for generating CI stimuli from digitized audio signals. For each participating subject, we used the Custom Sound™ software by Cochlear Limited to retrieve the processing parameters, i.e., the MAP, from his/her CI system. These parameters included the pulse rate and pulse width, the number of maxima, TL and CL for each channel, and others. The same parameter values were then used to generate the test stimuli for both ACE and ToneACE.

Procedures

Each subject was required to complete five test sessions on different days. A validation experiment was carried out in the first two sessions to compare the word and tone recognition performance in quiet with the subject's own CI processor (Session 1) and with the experimental platform (Session 2). This was to ensure that the software-implemented ACE strategy and the Laura 34 research processor were functioning properly, and that speech recognition performance with the research processor was comparable to that achieved with the subject's own CI processor. When listening via subjects' own CI processors, the subjects were seated one meter away in front of the loudspeaker (JBL LSR4326P) and acoustic stimuli were played via a loudspeaker at RMS level of 65 dBA in a quiet room. In the remaining three sessions, electrical stimuli were generated from noisy speech materials and delivered by direct stimulation via the Laura 34 experimental platform. Session 4 used test stimuli processed by the ToneACE method while Sessions 3 and 5 used test stimuli processed by the conventional ACE. Different SNR levels were used for different subjects, depending on the performance level of the subject with the clean speech. For subjects who could attain 70% correct or above on word identification, the SNR for tests with noisy speech was fixed at 0 dB. For those who scored below 70%, the SNR was set to 5 dB. For each subject, the SNR used was the same across processing conditions.

Each test session was 2 hours long and consisted of a familiarization phase and a test phase. In the familiarization phase, subjects received two blocks of training with 30 trials per block. For each trial, the subject was asked to identify the presented word from four candidates. Speech stimuli used for training were recorded from one male and one female speaker. The goal of the training session was to let the subjects be familiar with the stimuli under each particular test condition as well as the test procedure. This was done in all of the test sessions, so as to facilitate fair comparison among their results. After training, subjects received two testing blocks. Each block consisted of 120 trials, which covered all of the disyllabic words in the full test set. The first test block used stimuli spoken by a male, and the second test block used stimuli spoken by a female. The speakers were different to those in the familiarization phase. The listening test was designed as a four-alternative forced-choice (4AFC) disyllabic word identification task. For each trial, one of the words was

presented and the subject was asked to identify it from the four candidate words. The presentation order of the test words was randomized without repetition. The listening test was administered and controlled by computer software with a graphical user interface. The stimulus was presented only one time without repetition. The subject was encouraged to make a guess if he/she was not certain about the correct answer. After receiving the subject's response, the test system would proceed to the next trial. In the test phase, no feedback was provided. During practice, feedback was provided to the subject at the end of each trial by notifying them whether they had identified the test word correctly. In the case of incorrect identification, the computer interface showed the subject the correct answer and replayed the identified word and the presented word so as to help him/her distinguish them.

Data Analysis

The four candidate words used in each test trial included exactly one pair of intended contrasting tones and three different segmental structures as described in the test materials section. For each test condition, the average percent correct scores for word, tone, and segmental structure identification were computed for each subject and for the group. The chance levels for word, segmental structure, and tone identification are 25%, 37.5%, and 50%, respectively. Note that the ACE strategy was tested twice. The overall word [t(6)=1.92, p=0.103], segmental structure [t(6)=1.79, p=0.124], and tone [t(6)=2.0, p=0.088] identification performance was similar between the tests of session 3 and 5 with the ACE strategy. Thus, subsequent analyses were performed using the average scores of two tests for each subject.

A two-way repeated measures analysis of variance (ANOVA) was conducted to determine if there were statistically significance differences in word, segmental structure, and tone identification with the factors of testing conditions (Soundfield Quiet, ACE Quiet, ACE Noisy, and ToneACE Noisy) and speaker gender (Male, Female). There were no outliers based on the studentized residuals, and the residuals were normally distributed for each group, as assessed by Shapiro-Wilk test ($p > 0.05$). Paired-sample t-tests were performed for planned pairwise comparisons. Bonforreni correction was used to adjust for the alpha level ($\alpha = 0.017$) for three main effect contrasts (Soundfield Quiet vs. ACE Quiet, ACE Quiet vs. ACE Noisy, ACE Noisy vs. ToneACE Noisy). Alpha was also adjusted ($\alpha = 0.025$) for two simple main effect comparisons (ACE Quiet vs. ACE Noisy, ACE Noisy vs. ToneACE Noisy) at each level of gender (Maxell & Delaney, 1989).

Results

The main effects of test condition and speaker gender were significant for word [test condition: $F(3,18)=33.9$, $p<0.001$; gender: $F(1,6)=29.6$, $p=0.002$] and tone [test condition: $F(3,18)=17.7$, $p=0.001$; gender: $F(1,6)=58.6$, $p<0.001$] identification. For segmental structure identification, significant difference was found for the main effect of test condition [$F(3,18)=36.3$, $p<0.001$], but not for speaker gender [$F(1,6)=5.4$, $p=0.06$]. There was also significant interaction between test condition and speaker gender for all three measures [segmental structure: $F(3,18)=9.1$, $p=0.001$; word: $F(3,18)=10.1$, $p<0.001$; tone: $F(3,18)=5.1$, $p=0.01$].

Direct Stimulation vs. Sound Field with Conventional ACE Strategy

The average percent correct score on segmental structure, word, and tone identification in quiet attained by direct stimulation (Test Session 2) was compared with that by sound-field listening (Test Session 1). While the segmental structure score was slightly better with sound-field listening (88%) than with direction stimulation (85%) [$t(6)=3.5$, $p=0.013$], the word identification scores from listening with the Laura 34 research processor (mean = 76%) were very close to those of listening with their own CI systems (mean = 77%) [$t(6)=1.28$, $p=0.247$]. Performance on tone identification (85%) between the two listening modes was not significantly different [$t(6)=0.21$, $p=0.839$]. This indicates that the experimental platform was functioning properly and the subjects were able to adapt to the electrical stimuli generated by the MATLAB-implemented ACE strategy. The subjects expressed that the research processor provided the same sense of hearing as their own devices. The implication of these results is that if we observed significant difference in tone identification performance with our ToneACE method compared to conventional ACE with direct stimulation, it is not likely due to unknown factors related to direct stimulation.

Effects of Noise Using Conventional ACE Strategy

Each subject was tested with the ACE strategy in noise and in quiet. Overall word identification scores with clean speech ranged from 47.5% to 88.8% correct. For the two subjects with the lowest scores 47.5% (S05) and 64.6% (S02), the SNR used for subsequent test sessions was 5 dB. For the other five subjects, the SNR was fixed at 0 dB. Figure 8 shows the average word, segmental structure, and tone identification scores for clean speech (black bars) and noisy speech (gray bars) conditions with the ACE strategy. Results on different speaker genders are plotted separately (top: male voice; bottom: female voice). Averaged across speaker genders, performance levels of all three measures dropped significantly with the presence of noise [word: $t(6)=5.58$, $p=0.001$; segmental structure: $t(6)=5.24$, $p=0.002$; tone: $t(6)=5.03$, $p=0.002$]. The accuracy of word identification was 64% for noisy speech, as compared to 76% for clean speech. The accuracy of tone identification dropped from 85% to 77%, and that of segmental structure identification also from 85% to 77%. Although the noise effect was significant for both the male [segmental: $t(6)=4.6$, $p=0.004$; word: $t(6)=4.1$, $p=0.007$; tone: $t(6)=3.5$, $p=0.012$] and female [segmental: $t(6)=5.4$, $p=0.002$; word: $t(6)=5.7$, $p=0.001$; tone: $t(6)=4.7$, $p=0.003$] voices, the degradation was greater for the female voice than for the male voice by 7, 4, and 5 percentage points for word, segmental, and tone identification, respectively.

ToneACE vs. Conventional ACE Strategy

Figure 8 also shows the average word, segmental structure, and tone identification scores for noisy speech with the ToneACE method (white bars). Averaged across speaker genders, tone identification was higher with the ToneACE method than with the conventional ACE strategy by about 5 percentage points [$t(6)=2.9$, $p=0.028$]; however, the difference is not considered statistically significant after Bonferroni correction. Tone identification scores with the female voice were lower than those with the male voice by 10 percentage points. Paired-t tests showed that ToneACE improved tone identification for the female voice [$t(6)$

= 2.9, $p=0.025$], but not for the male voice [$t(6) = 1.3$, $p=0.236$]. The average improvement for the female voice was about 7 percentage points.

For segmental structure and word identification, there was no statistically significant difference between the ACE and ToneACE processing condition when averaged across speaker genders [segmental: $t(6)=0.5$, $p=0.60$; word: $t(6)=1.9$; $p=0.11$]. Post-doc analyses with paired-t tests also did not show significant difference between the two processing conditions for either the male or the female speaker ($p>0.05$).

Discussion

Benefit of Periodicity Enhancement for Pitch-Related Tasks

The present study showed that CI users' ability to perceive tone contrasts was reduced when speech was mixed with noise. This finding is consistent with results reported by previous studies on Mandarin tone recognition on normal-hearing listeners listening with CI simulation (Kong and Zeng, 2006) and on real CI users (Milczynski et al., 2012). The diminished temporal pitch performance could be the result of the reduction of modulation depth caused by the noise and the random jitters from noise envelope. Previous studies have shown the effect of modulation depth on temporal pitch perception. Geurts and Wouters (2001) reported that CI users' ability to discriminate pitch of amplitude-modulated stimuli decreased with decreasing modulation depth. In one extreme, when modulation depth was too small to be detected, the perceived pitch would correspond to that produced by pulse-train carrier instead of that produced by the rate of modulation (McKay et al., 1995), or if the carrier rate was too high, it would be out of the temporal pitch range for CI users.

Similar to previous studies (e.g., Green et al., 2004; Vandali et al., 2005; Laneau et al., 2006), the ToneACE method in the present study enhances the saliency of temporal periodicity cues by replacing the complex temporal periodicity cue of the original speech with an F0-controlled sinusoidal wave, increasing the modulation depth of the F0-related modulation and synchronizing the periodicity across channels. This enhancement method has been shown to improve pitch-related tasks, including perception of speech intonation (Green et al., 2005), as well as pitch and melody perception (Vandali et al., 2005; Laneau et al., 2006). For tone perception, Yuan et al. (2009) showed that for a group of normal-hearing listeners, the enhanced temporal periodicity cues in a four-channel noise-band vocoder improved Cantonese tone perception in noise. Similar to the patterns of results observed in the current study with real CI patients, Yuan et al. showed significantly poorer tone identification performance with a female voice than with a male voice and significantly improved tone identification with periodicity enhancement compared to without. Improvement for word and segmental structure identification was not evident in both studies. Recently, Milczynski et al. (2012) tested four Mandarin-speaking CI users with their F0mod strategy. They reported a significant improvement on Mandarin tone perception in noise while the performance of word recognition was unaffected by the tone-enhanced technique. Consistent with the findings of Milczynski et al., the ToneACE method used here also showed significantly better tone perception in noise for Cantonese CI users, and that word and segmental structure identification performance was similar between ACE and ToneACE.

Effect of F0 Range

Similar to previous findings on speech intonation (Green et al., 2005) and Mandarin tone perception (Milczynski et al., 2012), our results showed that tone perception was significantly better for a male voice than for a female voice for Cantonese tone perception. This is consistent with results from CI psychophysical studies, which showed poorer temporal pitch perception with higher amplitude-modulation frequencies for both normal-hearing (Burns and Viemesiter, 1976, 1981) and CI listeners (McKay et al., 1994; McDermott and McKay, 1997; Kong et al., 2009).

In our implementations of both the ACE and ToneACE methods, the outputs of individual frequency channels were obtained by summing up the power of respective frequency bins. The power-summation method was found to cause reduction of the modulation depth for higher modulation frequencies, i.e., 180 – 250 Hz (Milczynski et al., 2012). Since a larger modulation depth is required to detect the modulation at higher frequencies (Shannon, 1992; Kreft et al., 2010; Chatterjee and Oberzut, 2011), this could also explain the poorer tone recognition performance with the female voice than with the male voice. In the presence of additive noise, F0-related modulation depth was further decreased, making tone perception with the female voice even more difficult. It must be noted that recent Nucleus processors, including Freedom and CP810, compute channel outputs by vector summation, which may provide a more effective representation of F0 periodicity in mid-to-basal channels (Milczynski et al., 2012). However, the benefit of vector summation over power summation has not been confirmed.

The ToneACE strategy provided a greater improvement on female voice than the male voice. This suggests that the observed ToneACE benefit could be mainly attributed to the enhancement of F0 modulation in cases where it was poorly coded by ACE, such as high F0s. Based on this observation, we expected that tone perception improvement could be observed for male voice at lower SNRs when the modulation depth was reduced by the increase of noise level. In the present study, the average tone identification score with the ACE strategy was 84% for the male speaker, suggesting the possibility of ceiling effect. It is noted that Milczynski et al. (2012) reported a significant Mandarin tone perception benefit with the F0mod strategy for male voice, but not female voice. As discussed in their paper, the lack of significant tone perception benefit for the female voice was possibly due to the small sample size and to the large standard deviation of the difference in tone scores between ACE and the F0mod at each SNR.

Effect of Pulse Rate

As shown in Table 1, three of the subjects used a stimulation pulse rate of 500 Hz and the remaining four subjects had higher pulse rates from 900 Hz to 1800 Hz. At a pulse rate of 500 Hz, a relatively high F0 (i.e., female voice) may not be adequately represented. Table 3 compares the average tone identification score of the three low-pulse-rate subjects with that of the four high-pulse-rate subjects. Due to the small sample size, we could not determine if the effect of pulse rate is statistically significant between the two groups. However, there is a trend that the low-pulse-rate subjects showed (1) lower tone identification scores in quiet, (2) less pronounced effect of noise on tone identification, and (3) reduced periodicity

enhancement benefit with ToneACE for tone identification in noise compared to the high-pulse-rate group.

Limitations and Future Directions

In this study, we used ground-truth F0 information for periodicity enhancement. The F0 values were estimated from clean speech utterances using the auto-correlation algorithm in PRAAT, and then verified manually. In Milczynski et al. (2012), F0 information was also extracted from clean speech for investigating Mandarin tone perception in noise. Although both studies demonstrated the benefits of periodicity enhancement, a valid question to be asked is whether reliable F0 estimation can be done for noise-corrupted speech. Indeed, robust F0 estimation is a challenge for all existing periodicity enhancement algorithms (Vandali and van Hoesel, 2011; Milczynski et al., 2012). Recently, a number of robust techniques have been proposed to deal with low-SNR conditions (Gonzalez and Brookes, 2011; Jin and Wang, 2011; Chu and Alwan, 2012). In our laboratory, we investigated and evaluated a new algorithm of robust F0 estimation using exemplar-based models and spectral-temporal features (Huang and Lee, 2013). For speech noise at 5 dB and 0dB SNR, this algorithm achieved error rates of 6.6% and 14.0%, respectively, which are better than most of the existing algorithms (11.4% and 23.9% respectively). Given that the benefit of periodicity enhancement to tone perception hinges on the accuracy of pitch estimation, we continue to investigate methods for better F0 estimation of noisy speech. At the same time, it is yet to be determined to what extent the effectiveness of tone identification with ToneACE (or F0mod) would be negatively affected by imperfect F0 extraction in noise. On the other hand, improvement with ToneACE in the current study was observed only on tone identification, but not on word and segmental structure recognition. Further investigation is needed to show that the strategy is useful in real-life speech communication.

Acknowledgments

We are grateful to all the subjects for their participation in this study. This work was supported by the General Research Fund (Ref. CUHK 414108 and 413811) from Hong Kong Research Grants Council and a project grant from the Shun Hing Institute of Advanced Engineering, The Chinese University of Hong Kong. It was also partly supported by National Natural Science Foundation of China (11104316), Shanghai Natural Science Foundation (11ZR1446000), Open Research Fund Program of Key Laboratory of Speech and Hearing Sciences (East China Normal University), Ministry of Education. The last author was supported by the National Institute of Health (Grant no. R01-DC012300).

Glossary

ACE	Advanced Combination Encoder
CI	Cochlear implant
CIS	Continuous Interleaved Sampler
CL	Comfort Level
F0Sync	F0 Synchronized ACE
FFT	Fast Fourier transform
MEM	Multi-channel Envelope Modulation

RMS	Root-mean-square
TEPC	Temporal envelope and periodicity cues
ToneACE	Tone-enhanced Advanced Combination Encoder
TL	Threshold Level
SNR	Signal-to-noise ratio

References

- Boersma, P.; Weenink, D. Praat: doing phonetics by computer (Version 5.1.05) [Computer program]. 2009.
- Burns EM, Viemeister NF. Nonspectral pitch. *J Acoust Soc Am*. 1976; 60:863–869.
- Burns EM, Viemeister NF. Played-again SAM: Further observations on the pitch of amplitude-modulated noise. *J Acoust Soc Am*. 1981; 80:1655–1660.
- Chatterjee M, Oberzut C. Detection and rate discrimination of amplitude modulation in electrical hearing. *J Acoust Soc Am*. 2011; 130:1567–1580. [PubMed: 21895095]
- Chatterjee M, Peng SC. Processing F0 with cochlear implants: Modulation frequency discrimination and speech intonation recognition. *Hear Res*. 2008; 235:143–156. [PubMed: 18093766]
- Chen TH, Massaro DW. Seeing pitch: visual information for lexical tones of Mandarin-Chinese. *J Acoust Soc Am*. 2008; 123:2356–2366. [PubMed: 18397038]
- China Disabled Persons' Federation. Communique on major statistics of the second China national sample survey on disability. China Disabled Persons' Federation; 2008. Available from: http://www.cdpf.org.cn/english/contactus/content/2008-04/14/content_84989.htm
- Chu W, Alwan A. AFE: A statistical approach to F0 estimation under clean and noisy conditions. *IEEE Trans Audio, Speech and Language Processing*. 2012; 20(3):933–944.
- Ciocca V, Francis AL, Aisha R, Wong L. The perception of Cantonese lexical tones by early-deafened cochlear implantees. *J Acoust Soc Am*. 2002; 111:2250–2256. [PubMed: 12051445]
- Cochlear Limited. NIC v2 Software Interface Specification. 2006.
- Cochlear Limited. Clinical Guidance Document. 2009.
- Fu QJ, Nogaki G. Noise susceptibility of cochlear implant users: The role of spectral resolution and smearing. *J Assoc Res Otolaryngol*. 2005; 6:19–27. [PubMed: 15735937]
- Geurts L, Wouters J. Coding of the fundamental frequency in continuous interleaved sampling processors for cochlear implants. *J Acoust Soc Am*. 2001; 109:713–726. [PubMed: 11248975]
- Geurts L, Wouters J. Better place-coding of the fundamental frequency in cochlear implants. *J Acoust Soc Am*. 2004; 115:844–852. [PubMed: 15000196]
- Gfeller K, Turner C, Mehr M, Woodworth G, Fearn R, et al. Recognition of familiar melodies by adult cochlear implant recipients and normal-hearing adults. *Cochlear Implants Int*. 2002; 3:29–53. [PubMed: 18792110]
- Green T, Faulkner A, Rosen S. Enhancing temporal cues to voice pitch in continuous interleaved sampling cochlear implants. *J Acoust Soc Am*. 2004; 116:2298–2310. [PubMed: 15532661]
- Green T, Faulkner A, Rosen S, Macherey O. Enhancement of temporal periodicity cues in cochlear implants: Effects on prosodic perception and vowel identification. *J Acoust Soc Am*. 2005; 118:375–385. [PubMed: 16119358]
- Gonzalez S, Brookes M. A pitch estimation filter robust to high levels of noise (PEFAC). *Proc Eur Signal Process Conf*. 2011:451–455.
- Hilbert, D. *Grundzuege Einer Allgemeinen Theorie der Linearen Integralgleichungen*. Leipzig: Teubner; 1912.
- Houtsma AJM, Smurzynski J. Pitch identification and discrimination for complex tones with many harmonics. *J Acoust Soc Am*. 1990; 87:304–310.

- Huang F, Lee T. Pitch Estimation in Noisy Speech Using Accumulated Peak Spectrum and Sparse Estimation Technique. *IEEE Trans Audio, Speech and Language Processing*. 2013; 21:99–109.
- Jin Z, Wang D. MM-based multipitch tracking for noisy and reverberant speech. *IEEE Trans Audio, Speech and Language Processing*. 2011; 19(5):1091–1102.
- Kiefer J, Hohl S, Stürzebecher E, Pfennigdorff T, Gstöttner W. Comparison of Speech Recognition with Different Speech Coding Strategies (SPEAK, CIS, and ACE) and Their Relationship to Telemetric Measures of Compound Action Potentials in the Nucleus CI 24M Cochlear Implant System: Comparación del reconocimiento del len. *Int J Audiol*. 2001; 40:32–42.
- Kong YY, Deeks JM, Axon PR, Carlyon RP. Limits of temporal pitch in cochlear implants. *J Acoust Soc Am*. 2009; 125:1649–1657. [PubMed: 19275322]
- Kong YY, Zeng FG. Temporal and spectral cues in Mandarin tone recognition. *J Acoust Soc Am*. 2006; 120:2830–2840. [PubMed: 17139741]
- Kong YY, Cruz R, Jones JA, Zeng FG. Music perception with temporal cues in acoustic and electric hearing. *Ear Hear*. 2004; 25:173–185. [PubMed: 15064662]
- Kreft HA, Oxenham AJ, Nelson DA. Modulation rate discrimination using half-wave rectified and sinusoidally amplitude modulated stimuli in cochlear-implant users. *J Acoust Soc Am*. 2010; 127:656–659. [PubMed: 20136187]
- Laneau, J. Doctoral dissertation, Ph D dissertation. Katholieke Universiteit Leuven, Faculteit Toegepaste Wetenschappen; Leuven, Belgium: 2005. When the deaf listen to music Pitch perception with cochlear implants.
- Laneau J, Wouters J, Moonen M. Improved music perception with explicit pitch coding in cochlear implants. *Audiol Neurotol*. 2006; 11:38–52.
- Lee KYS, van Hasselt CA, Chiu SN, Cheung DMC. Cantonese tone perception ability of cochlear implant children in comparison with normal-hearing children. *Int J Pediatr Otorhinolaryngol*. 2002; 63:137–147. [PubMed: 11955605]
- Li T, Fu QJ. Effects of spectral shifting on speech perception in noise. *Hear Res*. 2010; 270:81–88. [PubMed: 20868733]
- Licklider, JCR. Three auditory theories. In: Sigmund, K., editor. *Psychology: A Study of a Science*. New York: McGraw-Hill; 1959. p. 41-144.
- Licklider JCR. A duplex theory of pitch perception. *Experientia*. 1951; 7:128–134. [PubMed: 14831572]
- Linguistic Society of Hong Kong. Hong Kong Jyut Ping Character Table. Hong Kong: Linguistic Society of Hong Kong Press; 1997.
- Maxell, SE.; Delaney, HD. *Designing Experiments and Analyzing Data – A Model Comparison Perspective*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc; 1989.
- McDermott HJ, McKay CM. Musical pitch perception with electrical stimulation of the cochlea. *J Acoust Soc Am*. 1997; 101:1622–1631. [PubMed: 9069629]
- McKay CM, McDermott HJ, Clark GM. Pitch matching of amplitude-modulated current pulse trains by cochlear implantees: the effect of modulation depth. *J Acoust Soc Am*. 1995; 97:1777–1785. [PubMed: 7699159]
- McKay CM, McDermott HJ, Clark GM. Pitch percepts associated with amplitude-modulated current pulse trains in cochlear implantees. *J Acoust Soc Am*. 1994; 96:2664–2673. [PubMed: 7983272]
- Meddis R, O'Mard L. A unitary model of pitch perception. *J Acoust Soc Am*. 1997; 102:1811–1820. [PubMed: 9301058]
- Milczynski M, Chang JE, Wouters J, van Wieringen A. Perception of Mandarin Chinese with cochlear implants using enhanced temporal pitch cues. *Hear Res*. 2012; 285:1–12. [PubMed: 22361414]
- Milczynski M, Wouters J, van Wieringen A. Improved fundamental frequency coding in cochlear implant signal processing. *J Acoust Soc Am*. 2009; 125:2260–2271. [PubMed: 19354401]
- National Institute on Deafness and Other Communication Disorders (NIDCD). NIDCD fact sheet: cochlear implants. Bethesda, MD: NIDCD Information Clearinghouse; 2009. Available from <http://www.nidcd.nih.gov/staticresources/health/hearing/FactSheetCochlearImplant.pdf>

- Nelson DA, Kreft HA, Anderson ES, Donaldson GS. Spatial tuning curves from apical, middle, and basal electrodes in cochlear implant users. *J Acoust Soc Am*. 2011; 129:3916–3933. [PubMed: 21682414]
- Nimmons GL, Kang RS, Drennan WR, Longnion J, Ruffin C, et al. Clinical assessment of music perception in cochlear implant listeners. *Otol Neurotol*. 2008; 29:149–155. [PubMed: 18309572]
- Oppenheim, AV.; Schafer, RW.; Buck, JR. Discrete-time signal processing. 2. New Jersey: Prentice-Hall; 1999.
- Peng SC, Tomblin JB, Turner CW. Production and perception of speech intonation in pediatric cochlear implant recipients and individuals with normal hearing. *Ear Hear*. 2008; 29:336–351. [PubMed: 18344873]
- Plomp R. Pitch of complex tones. *J Acoust Soc Am*. 1967; 41:1526–1533. [PubMed: 6075560]
- Qin MK, Oxenham AJ. Effects of envelope-vocoder processing on F0 discrimination and concurrent-vowel identification. *Ear Hear*. 2005; 26:451–460. [PubMed: 16230895]
- Ritsma R. Frequencies dominant in the perception of the pitch of complex sounds. *J Acoust Soc Am*. 1967; 42:191–198. [PubMed: 6052077]
- Rosen S. Temporal information in speech: acoustic, auditory and linguistic aspects. *Philos Trans R Soc London, Ser B*. 1992; 336:367–373. [PubMed: 1354376]
- Schouten JF. The residue and the mechanism of hearing. *Proc Kon Akad Wetenschap*. 1940; 43:991–999.
- Shannon RV. Temporal modulation transfer functions in patients with cochlear implants. *J Acoust Soc Am*. 1992; 91:2156–2164. [PubMed: 1597606]
- Smith ZM, Delgutte B, Oxenham AJ. Chimaeric sounds reveal dichotomies in auditory perception. *Nature*. 2002; 416:87–90. [PubMed: 11882898]
- Stickney GS, Zeng F-G, Litovsky R, Assmann P. Cochlear implant speech recognition with speech maskers. *J Acoust Soc Am*. 2004; 11(6):1081–1091. [PubMed: 15376674]
- Swanson, B.; Mauch, H. Nucleus MATLAB Toolbox 4.20 [computer software and user manual]. Cochlear Limited; 2006.
- Vandali AE, van Hoesel RJM. Development of a temporal fundamental frequency coding strategy for cochlear implants. *J Acoust Soc Am*. 2011; 129:4023–4036. [PubMed: 21682423]
- Vandali AE, van Hoesel RJM. Enhancement of temporal cues to pitch in cochlear implants: Effects on pitch ranking. *J Acoust Soc Am*. 2012; 132:392–402. [PubMed: 22779486]
- Vandali AE, Sucher C, Tsang DJ, McKay CM, Chew JWD, et al. Pitch ranking ability of cochlear implant recipients: A comparison of sound-processing strategies. *J Acoust Soc Am*. 2005; 117:3126–3138. [PubMed: 15957780]
- Vandali AE, Whitford LA, Plant K, Clark GM. Speech perception as a function of electrical stimulation rate: Using the Nucleus 24 cochlear implant system. *Ear Hear*. 2000; 21:608–624. [PubMed: 11132787]
- Wei CG, Cao K, Zeng FG. Mandarin tone recognition in cochlear-implant subjects. *Hear Res*. 2004; 197:87–95. [PubMed: 15504607]
- Wilson BS, Finley CC, Lawson DT, Wolford RD, Eddington DK, et al. Better speech recognition with cochlear implants. *Nature*. 1991; 352:236–238. [PubMed: 1857418]
- Wong LLN, Vandali AE, Ciocca V, Luk B, Ip VWK, et al. New cochlear implant coding strategy for tonal language speakers. *Int J Audiol*. 2008; 47:337–347. [PubMed: 18569106]
- Xu L, Pflingst BE. Relative importance of temporal envelope and fine structure in lexical-tone perception. *J Acoust Soc Am*. 2003; 114:3024–3027. [PubMed: 14714781]
- Xu, L.; Zhou, N. Tonal Languages and Cochlear Implants. In: Zeng, F-G.; Popper, AN.; Fay, RR., editors. *Auditory Prostheses: New Horizons*, Springer Handbook of Auditory Research. Vol. 39. New York: Springer Science + Business Media, LLC; 2012. p. 341–364.
- Yuan M, Lee T, Yuen KCP, Soli SD, van Hasselt CA, et al. Cantonese tone recognition with enhanced temporal periodicity cues. *J Acoust Soc Am*. 2009; 126:327–337. [PubMed: 19603889]
- Yuen KCP, Pang KW, Tong MCF, van Hasselt CA, Yuan M, et al. Development of the computerized Cantonese Disyllabic Lexical Tone Identification Test in Noise (CANDILET-N). *Cochlear Implants Int*. 2009; 10:130–137. [PubMed: 19195002]

Zeng FG. Temporal pitch in electric hearing. *Hear Res.* 2002; 174:101–106. [PubMed: 12433401]

Figure 1.

Schematic descriptions of Mandarin and Cantonese tones. The numeral labels of Cantonese tones followed the Jyut Ping system of the Linguistic Society of Hong Kong (LSHK 1997).

Figure 2.

Temporal envelopes of Mandarin and Cantonese syllable. Top panel shows the temporal envelopes of Mandarin syllable /ma/ carrying four different tones and the bottom panel shows temporal envelopes of Cantonese syllable /ji/ carrying six different tones.

Figure 3.

An example of four disyllabic words displayed on a computer screen for a given test trial. Two of the words differ in tone contrast /ging1 lik6/ (經歷) vs. /ging2 lik6/ (警力). The other two words /gung1 lik6/ (功力) and /geng2 lik6/ (頸力) have the same intended tone contrast (i.e., tone 1 vs. tone 2), but with different segmental compositions.

Figure 4.
Implementation of the conventional ACE strategy.

Figure 5.
Implementation of ToneACE.

Figure 6.

Visual comparison of the original TEPC $p_i[n]$ and enhanced TEPC $p'_i[n]$ at one of the stimulation channels. The speech segment contains the Cantonese test word syllable /ging1 lik6/ with additive speech-shaped noise at 0 dB SNR. The thick line in the top and bottom panel is the very slow varying envelope $e_i[n]$ of the speech signal for that channel.

Figure 7.

Electrodegrams for the Cantonese syllable /gei1/ spoken by a male native speaker at 0 dB SNR generated by ACE and ToneACE. The y-axis of the electrodegrams shows the indices of the electrodes ordered from apex (electrode 22) to base (electrode 1).

Figure 8. Comparison of word, segmental structure, and tone identification performance with different listening and processing conditions. All listening tests were performed with a direct stimulation mode. Results for the male and female voice are shown in separate panels.

Table 1

Subject Demographics

Subject	Gender	Age	Etiology	Duration of hearing loss (yrs)	Implant experience (yrs)	Speech processor	Pulse rate (Hz)	Number of maxima
S01	F	57	Unknown	20	5	Freedom	500	10
S02	F	59	Otosclerosis	3	10	ESPrit 3G	1200	10
S03	F	43	Chemotherapy	0.5	0.5	CP 810	900	12
S04	M	55	Unknown	3	13	SPrint	1800	8
S05	M	68	Chemotherapy	6	6	ESPrit 3G	500	9
S06	M	53	Unknown	1.5	13	Freedom	500	9
S07	M	60	Unknown	2	8	SPrint	900	11

Table 2

List of Testing Conditions.

Session	Testing mode	Speech processor	Speech Materials	Processing strategy
1	Sound Field	clinical processor	Clean	ACE
2	Direct Stimulation	Laura 34	Clean	ACE
3	Direct Stimulation	Laura 34	Noisy	ACE
4	Direct Stimulation	Laura 34	Noisy	ToneACE
5	Direct Stimulation	Laura 34	Noisy	ACE

Table 3

Average Tone Identification Scores for Low-Pulse-Rate and High-Pulse-Rate Subjects on Female Voice.

Test condition	Low-pulse-rate subjects	High-pulse-rate subjects
Clean	76.1%	84.6%
Noisy with ACE	70.6%	70.9%
Noisy with ToneACE	73.1%	82.1%