

# Simulated Phase-Locking Stimulation: An Improved Speech Processing Strategy for Cochlear Implants

Jing Chen<sup>a</sup> Xihong Wu<sup>a,b</sup> Liang Li<sup>a,c</sup> Huisheng Chi<sup>a,b</sup>

<sup>a</sup>National Key Laboratory on Machine Perception, Speech and Hearing Research Center, and <sup>b</sup>Department of Psychology, Peking University, Beijing, PR China; <sup>c</sup>Department of Psychology, Centre for Research on Biological Communication Systems, University of Toronto at Mississauga, Mississauga, Ont., Canada

## Key Words

Speech processing strategy · Cochlear implant · Phase information · Chinese speech

## Abstract

The continuous interleaved sampling (CIS) speech-process-

## Introduction

Cochlear implant (CI) devices have been applied successfully to help profoundly deaf patients achieve hearing through electrical stimulation of the auditory nerve with fine electrodes inserted into the scala tympani of the cochlea [1]. The performance of listeners using CI devices depends largely on the signal processor transforming speech signals to electrical stimuli. Several signal-processing techniques have been developed over the past 30 years, and have been classified into 2 major types: waveform representation and feature extraction. As a typical waveform representation approach, the continuous interleaved sampling (CIS) strategy developed by researchers at the Research Triangle Institute shows a high level of speech recognition for the CI users speaking monotonal languages, such as English and German [2–4].

However, it has been reported that CI users who speak Chinese have poor identification of vowels and consonants [5, 6]. Chinese is a tonal language, which has 4 tonal patterns as defined by the fundamental frequency (F0) of voiced speech. For example, changing the tone in the syllable ‘ma’ from flat to rising, or to falling and rising, or to falling, changes the meaning of the word. Using the CIS strategy, Xu et al. [7] studied how signal-processing parameters, such as the low-pass cutoff frequency for

extracting amplitude envelopes and the number of channels of the band-pass filter bank, affect tonal recognition. The results of their studies show that recognition of the 4 Mandarin tonal patterns depends on both the number of channels and the low-pass cutoff frequency, and temporal cues can compensate for diminished spectral cues in tone recognition and vice versa. In addition, the importance of pitch and periodicity information in Chinese speech recognition have also been confirmed in the study by Fu et al. [8], in which 3 carrier band conditions were tested, including noise-band carrier for all speech segments, pulse train carriers for the voiced speech segment whose rate followed the F0 of the speech signals, and fixed-rate pulse train carriers for voiced speech segments. The results show that the F0-controlled pulse train carriers produce the best performance, indicating the need to provide adequate amounts of both pitch and periodicity information to Chinese-speaking CI patients.

Although some CI users perform well in speech recognition as normal listeners in a quiet environment, they have considerable difficulties in performance when maskers, especially fluctuating maskers, are presented [9]. F0 information has long been thought to play an important role in perceptually segregating sound sources [10]. A reduction in F0 cues produced by cochlear-implant processing leads to difficulty in segregating different sources. Moreover, fine structure information is also important for sound localization and pitch perception [11]. So, it is important to study how to convey more fine structure information of the speech signal to CI users.

Although in some CI strategies, such as MPEAK (multi-peak), F0, the first formant, and the second formant are extracted and used to modulate the electrical pulse's firing, errors are induced in formant extractions, especially in the situations where the speech signals are embedded in noise [1]. According to the CIS strategy the envelope information of band-pass filtered speech sounds are extracted and used to modulate the amplitude of electrical stimulation pulses of implanted electrodes without preserving the phase information in speech sounds. Since the phase information is potentially useful for improving CI listeners' speech perception [12], the present study proposes a new CI speech-processing strategy, the simulated phase-locking stimulation (SPLS) strategy, which preserves part of phase information in original speech and would be useful for upgrading the function of a CI device by introducing phase-related modulation of stimulation-pulse intervals. To experimentally evaluate the

efficacy of the SPLS strategy in processing Mandarin Chinese speech, we presented the acoustic stimulation of the SPLS strategy to normal-hearing Chinese listeners under either noise-masking or competing-speech-masking conditions.

## Methods

### *Simulated Phase-Locked Stimulation Strategy*

Figure 1 illustrates how the SPLS strategy extracts envelopes of band-pass filtered signals and uses phase information to modulate pulse rates [1, 6]. A signal is pre-emphasized first and then decomposed into multiple frequency bands by a bank of band-pass filters. Because in the present study the filter-bank should not distort phases of input signal components, the zero-phase transfer function is used in the stage of band-pass filtering [13]. After that, the signal in each band goes through 2 signal pathways: envelope extraction and phase extraction. To extract envelope information, the filtered signal is processed by the Hilbert transform and the extracted envelope is then logarithmically compressed to an acceptable dynamic range for CI. The compressed envelope will be used to modulate the amplitude of pulse trains that are interleaved among electrodes. To extract phase information, the 'zero-crossing detection' process was used to record every zero-crossing time of the narrow-band signal in each band. The phase information will be used to decide the firing time of pulse trains.

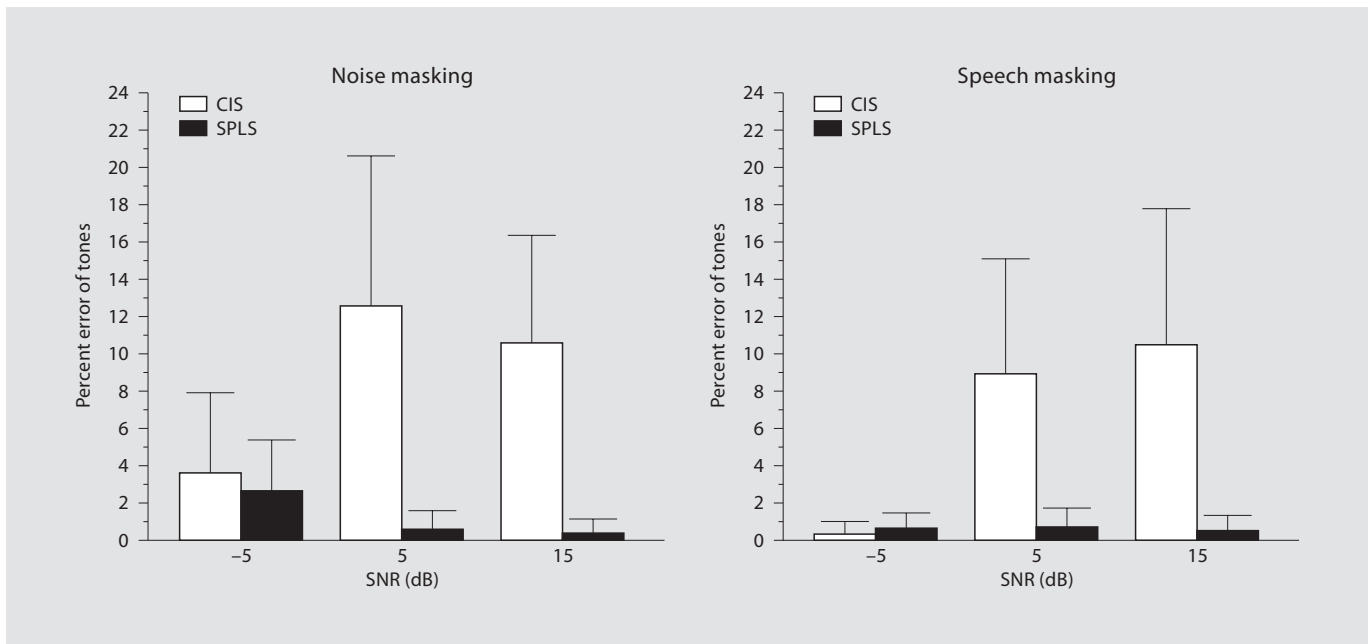
The pulse-firing strategy of SPLS simulates the neural mechanism of human hearing. In the human auditory system, the nerve firings occur at roughly the same phase of the waveform each time. However, there is also a difference between low and high frequencies. In detail, a single auditory nerve fiber fires on every cycle of tone stimulus in the low-frequency range and does not necessarily fire on every cycle of tone stimulus in the high-frequency range. In SPLS, the electrical stimulation pulses of each channel occur at the zero-phase of the signal in the corresponding channel. For a given channel whose center frequency is below 1,200 Hz, pulses fire at every zero-crossing time detected from the band-pass filtering signal. Otherwise, pulses fire once every  $\lceil f/1,200 \rceil$  zero-crossing times, where  $f$  is the center frequency and  $\lceil \cdot \rceil$  means the smallest integer bigger than  $f/1,200$ . The amplitude of the pulse is modulated by the extracted envelope.

For the CIS strategy the periods between pulses in each channel are fixed and simultaneous firing across channels can be avoided. However, for the SPLS strategy, the pulse rate in each channel is changed according to phase information, and simultaneous firing between 2 adjacent channels will happen. So, we measured the possibility of simultaneous firing between 2 adjacent channels on a 49-second piece of sound (including male or female English speech, Chinese speech, and a piece of music), which was processed by the SPLS strategy with 8 channels. When 2 pulses of 2 adjacent channels, respectively, fired at the same time, this firing was counted as a simultaneous firing. The final percent of simultaneous firing was 1.9%, which was too small to use additional inhibitory procedures.

#### *Acoustic Simulation*

Previous studies have confirmed that examination of normal-hearing listeners' responses to acoustic simulation of a CI processing strategy is useful for evaluating these strategies [14]. Thus,

ant is quarrelling with a bag, whose direct Chinese translation sounds like: ‘Yi1 ma3 zi3 zheng4 zai4 xuan1 nao4 zi1 ge1 shu1 bao1’, all the 3 underlined words are the key words [18]. Target speech stimuli were spoken by a young female speaker, and tested in a quiet environment or 1 of the 2 masking conditions,



**Fig. 3.** Mean percent-error in recognition of tones across 12 subjects as a function of SNR for each of the 2 processing strategies under 2 masking conditions: steady-spectrum-noise masking and speech masking. The error bars indicate the SD of the mean.

ANOVA analysis shows that the difference is significant,  $F(1, 11) = 55.288$ ,  $MSE = 372.09$ ,  $p = 0.000$ .

As shown in figure 2, under masking conditions speech recognition increased with the increase of the SNR in all conditions, and the recognition of the target speech processed by SPLS was much larger than that processed by CIS in both noise and speech-masking conditions. The main effect of SNR was significant,  $F(1, 11) = 448.33$ ,  $MSE = 4.642$ ,  $p = 0.000$ , the main effect of processing strategy was significant,  $F(1, 11) = 656.473$ ,  $MSE = 4.821$ ,  $p = 0.000$ , and the main effect of masking type was significant,  $F(1, 11) = 102.406$ ,  $MSE = 0.471$ ,  $p = 0.000$ .

To examine whether the SPLS strategy was also beneficial to recognition of tones, we analyzed the ‘tone error’ in sentence repeating across 12 subjects. The percent error in recognizing tones was defined as the percentage of the number of Chinese characters whose syllable was correctly recognized but whose tone was not correctly pronounced out of the number of 108, which was the total number of keyword characters in each list. Under the quiet condition, the mean percent-error in recognition of tones was 0.29% for the SPLS strategy and 8.17% for the CIS strategy. Under masking

conditions, the percent error in recognition of tones was much less for the SPLS strategy than for the CIS strategy. Under the low SNR condition (SNR = 5 dB), the difference between the SPLS strategy and the CIS strategy was not significant. However, when the SNR was increased to 5 or 15 dB, the percent error in recognition of tones was decreased more for the SPLS strategy than for the CIS strategy (fig. 3).

## Discussion

As pointed out by Fu et al. [19], there are additional needs for developing speech-processing strategies to specifically improve functions of cochlear implant devices for recognizing tonal languages, such as Chinese. Phase information is presented in speech for normal listeners, and is important not only for sound localization, but also for signal recognition in noise [12]. In the present study, adding phase information with the SPLS method into target speech remarkably improved listeners’ recognition performance in quiet. More importantly, additional phase information presented in target speech released the speech from noise and speech maskers.



- 11 Smith ZM, Delgutte B, Oxenham AJ: Chimeric sounds reveal dichotomies in auditory perception. *Nature* 2002;416:87-90.
- 12 Clopton BM, Spelman FA: Technology and the future of cochlear implants. *Ann Otol Rhinol Laryngol Suppl* 2003;191:26-32.
- 13 Mitra SK: *Digital Signal Processing: A Computer-Based Approach*. New York, McGraw-Hill, 2002.
- 14 Roggero MA, Robles L, Rich NC, Costalupes JA: Basilar membrane motion and spike initiation in the cochlear nerve; in Moore BCJ, Patterson RD (eds): *Auditory Frequency Selectivity*. New York, Plenum, 1986.
- 15 Glasberg BR, Moore BCJ: Derivation of auditory filter shapes from notched-noise data. *Hear Res* 1990;47:103-138.
- 16 Helfer KS: Auditory and auditory-visual perception of clear and conversational speech. *J Sp Lan Hear Res* 1997;40:432-443.
- 17 Li L, Daneman M, Qi JG, Schneider BA: Does the information content of an irrelevant source differentially affect speech recognition in younger and older adults? *J Exp Psychol Hum Percept Perform* 2004;30:1077-1091.
- 18 Wu XH, Wang C, Chen J, Qu HW, Li WR, Wu YH, Schneider BA, Li L: The effect of perceived spatial separation on informational masking of Chinese speech. *Hear Res* 2005;199:1-10.
- 19 Fu QJ, Hsu CJ, Horng MJ: Effects of speech processing strategies on Chinese tone recognition by Nucleus-24 cochlear implant users. *Ear Hear* 2004;25:501-508.
- 20 Lan N, Nie KB, Gao SK, Zeng FG: A novel speech processing strategy incorporating tonal information for cochlear implants. *IEEE Trans Biomed Eng* 2004;51:752-760.