Improved Mandarin Chinese Perception for Cochlear Implants Using an Auditory Model based on Stochastic Stimulation

Xiao-Jun Zhang, Yi Cao, Wen-Ye Sun, He-Ming Zhao, Di Wu, and Zhi Tao

***Abstract*—In this study, we propose an auditory model based on stochastic stimulation to enhance the low frequency (LF) information for a cochlear implant, which improves the perception of Mandarin Chinese. For LF channels, a model of the peripheral auditory system is used to generate the stochastic impulse, which is closer to the physiological activity of auditory neural firing. For high frequency channels, interleaved impulses are generated when no impulses appear in the LF channels. Compared with previous methods, such as the continuous interleaved sampling and zerocrossing-based stimulation strategies, our method makes three main contributions: i) both envelope and impulse intervals are employed to transmit speech information; ii) Mandarin Chinese perception is improved in noisy and quiet environments; and iii)the robustness of the proposed strategy is higher and even valid in noisy situations.**

***Index Terms*—Cochlear implant, Continuous interleaved sampling, Mandarin Chinese tones, Neural firing, Peripheral auditory system**

I. Introduction

Cochlear implants electrically stimulate the auditory nerve to restore hearing in profoundly deaf people. Remarkable progress has been achieved in the design of the cochlear implant system, but many improvements are still needed. Indeed, the speech identification performance is still quite poor for cochlear implant users who speak tonal languages, such as Mandarin Chinese [1].

In tonal languages, the range of the pitch of a vowel sound defines the meaning of the word in which it occurs. In Mandarin, the pitch trajectory of a vowel has four different variations called tones, which indicate different lexical meanings. The four tones are flat (-), rising (/), falling-rising (\/), and falling (\). A single consonant-vowel combination can have four lexical meanings depending on the tone. Therefore, different tones must be classified by listeners to facilitate accurate speech perception.

According to Smith, the poorness of perception for tonal languages is due mainly to the fact that the spectral information presented by modern cochlear implant speech processing strategies is insufficient for reliable tone recognition [2].Thus, delivering more spectral information to cochlear implant recipients should allow more accurate tonal language speech perception.

All current cochlear implants, except those that deliver analog waveforms [3], employ speech processing strategies based on the extraction and representation of temporal envelope information, such as the continuous interleaved sampling (CIS) strategy [4] and NofM strategy [5]. The impulse stimulation is applied at a constant rate in these strategies. In 2005, Nie *et al.* proposed a frequency amplitude-modulation encoding strategy, which encodes the temporal envelope, as well as the frequency modulation information using a slow-varying signal extracted from the fine structure [6]. Recently, it was found that the rate of impulse stimulation can change the perceived pitch [7], and these so-called “variable stimulation rates” can be used with pulse stimulation to include more spectral information [8–10]. Recently, Chen and Zhang explored the possibility of applying zerocrossing-based non-uniform stimulation to deliver low frequency (LF) information, and they found that the zerocrossing rate can also convey Mandarin tone information efficiently [11]. However, all of these strategies for cochlear implants only provide a very crude approximation of normal cochlear processing.

In this study, we proposed a stimulation algorithm based on a computational model of the peripheral auditory system to generate the stochastic impulse trains of the auditory nerves. Earlier studies suggest that the information in LF channels carries more perceptual cues to facilitate Mandarin tone identification [12]. Therefore, we applied our proposed stimulation algorithm to the LF channels and we also conducted acoustic simulation experiments to evaluate the performance of the proposed strategy in Mandarin speech perception based on comparisons with the CIS and zerocrossing-based strategies.

Manuscript received May 20, 2015; revised November 23, 2015. This work was supported in part by National Natural Science Foundation of China under Grant No.61271359 and No.61372146.

X.J. Zhang is with College of Physics, Optoelectronics and Energy, Soochow University, Suzhou 215006, China.

Y. Cao is with School of Electronic and Information Engineering, Soochow University, Suzhou 215006, China.

W.Y. Sun is with College of Physics, Optoelectronics and Energy, Soochow University, Suzhou 215006, China.

H.M. Zhao is with School of Electronic and Information Engineering, Soochow University, Suzhou 215006, China.

D. Wu is with College of Physics, Optoelectronics and Energy, Soochow University, Suzhou 215006, China.

Z. Tao is with College of Physics, Optoelectronics and Energy, Soochow University, Suzhou 215006, China (tel.:+86051267870162; fax:+86051267870271;e-mail:taoz@suda.edu.cn).

(Advance online publication: 29 February 2016)

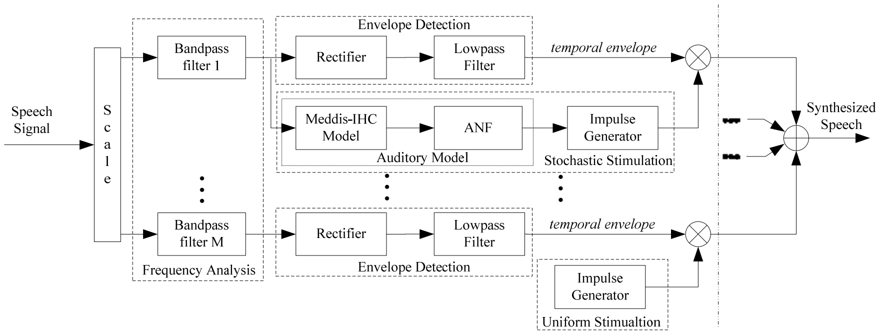

Fig.1. Block diagram of the proposed strategy.

II. Methodology

A. General Structure

In this study, we proposed an auditory model based on a stochastic stimulation algorithm to deliver Mandarin Chinese information. A block diagram of the proposed strategy is illustrated in Fig.1. The structure was originally motivated by the place coding (tonality) of the basilar membrane [13]. First, the speech signal was scaled so it was equivalent to the L-level of sound pressure in dB [14]. Second, the scaled speech signal was decomposed into M channels using a bank of band-pass filters. Third, the relative strengths of multiple channels were obtained by envelope detection and the envelopes of the sub-bands were then utilized to modulate the amplitudes of the stimulus impulses.

In our study, the stimulus impulses were produced in the following manner. In the LF channels (i.e., those covering frequencies below 1000 Hz), an auditory model of the peripheral auditory system, which comprises the inner hair cell (IHC) model and the auditory nerve fiber (ANF) model, was employed to determine the time required to evoke an impulse. By contrast, in the remaining high frequency (HF) channels, the impulses utilized the traditional uniform stimulation algorithm to evoke impulses. The impulse train generated for each channel was amplitude-modulated by the temporal envelope to produce a specific reconstructed signal. Finally, the synthesized speech was produced by summing the outputs from all the channels.

B. Stochastic Impulse Generation

As shown in Fig.1, an auditory model of the peripheral auditory system, which comprises the Meddis IHC model and the ANF model, was employed to generate the impulse. Compared with other hair cell models, the Meddis IHC model is described using simple mathematical expressions and it can reflect the activity of the auditory system better [17]. Therefore, in this study, we employed the Meddis IHC model, which comprises three reservoirs and one factory. The three reservoirs were called the free transmitter pool, synaptic cleft, and reprocessing store. Let q(t), c(t), and w(t) be the amounts of transmitter in the free transmitter pool, the synaptic cleft, and the reprocessing store, respectively. We refer to the factory as the "transmitter production replenisher."Figure 2 shows the data flow in the Meddis IHC model. Three compartments transfer the transmitter in a re-uptake and re-synthesis process loop. The factory delivers transmitters to the pool and the pool releases transmitters to the cleft. Some transmitters are lost by diffusion and some are taken back to the store. The remaining transmitters determine the firing rates of auditory nerve cells. The procedure of the Meddis IHC model can be described by the following four equations:

$$k(t)=\begin{cases} gdt[s(t)+A]/[s(t)+A+B], & [s(t)+A]\geq 0 \\ 0 & [s(t)+A]<0 \end{cases} \quad (1)$$

$$\frac{dq}{dt}=y[M-q(t)]+xw(t)-k(t)q(t) \quad (2)$$

$$\frac{dc}{dt}=k(t)q(t)-lc(t)-rc(t) \quad (3)$$

$$\frac{dw}{dt}=rc(t)-xw(t), \quad (4)$$

where y, x, l, r, g, A, B, and M are the parameters of the model, and their value sin the implementation of the IHC block are: y = 5.05,x = 66.31,l = 2500, r = 6580,g = 2000, A = 5,B = 300,and M = 1 [12].

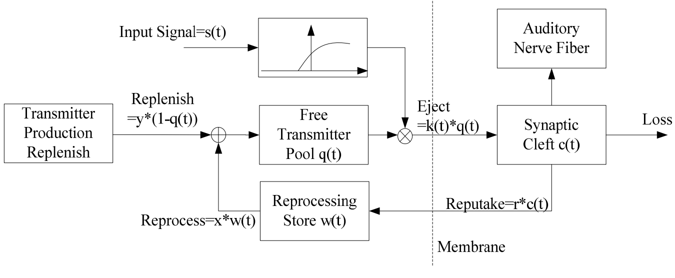

Fig.2. Data flow in the Meddis IHC model.

The remaining level of transmitter in the cleft $c(t)\cdot dt$ determines the probability of the afferent auditory nerve firing:

$$p(t) = hc(t)dt, \tag{5}$$

where the parameter h = 50000 [16]. The membrane potential of the ANF increases when the transmitter is released into the synapse between the IHC and the auditory nerve. The ANF generates a neural impulse when the probability exceeds a certain threshold. Thus, the block of decision logic can be used to determine whether or not the impulse is produced. An impulse occurs if p(t) is higher than a specific stochastic value [16].

$$spike_i(t) = \begin{cases} 1 & (p(t) \geq random(0...1)) \;\; and \;\; (\Delta t \geq 1ms) \\ 0 & others \end{cases}, \qquad i = 1,...,20 \tag{6}$$

The auditory nerve has a refractory period with a certain length (about 1ms), during which it is incapable of acting after the neural impulse is generated [17]. Normally, an IHC is connected with several ANFs. In our study, we connected 20 ANFs with each IHC. Finally, the neural impulses are generated after a logical operation OR, which is carried out on the respective outputs of the 20 ANFs associated with an IHC.

$$spike(t) = \mathrm{OR}(spike_1(t), spike_2(t), ..., spike_{20}(t)) \tag{7}$$

C. Speech Synthesis Model

In a similar manner to the traditional CIS strategy, to simulate the HF channels, the carrier signal in each channel uses the sinusoid with the same central frequency as the band-pass filter in that channel, which is then amplitude-modulated by the temporal envelope ai(t) in the same channel:

$$recon_i(t) = a_i(t)\cdot\cos(f_i\cdot t + \theta), \tag{8}$$

where θ is the initial phase of the signal. Its influence is neglected in the current study, i.e.,θ is set to 0.

For the LF channels, the impulse train at channel i is used to produce a continuous signal, $C_i(t)$. First, the k_{th} impulse generated at time t_k is determined from the speech signal. Then, $C_i(t)$ is generated using a train of high-rate sinusoidal pulses[18], $con(t)$, which are located at the impulse times, as follows.

$$C_i(t) = \sum_k con(t - t_k) \tag{9}$$

Ci(t) is then amplitude-modulated by its temporal envelope *ai(t)*to produce the band-specific decomposition output.

$$recon_i(t) = a_i(t)\cdot C_i(t) \tag{10}$$

Finally, the synthesized speech is obtained by summing the decomposition outputs from all the channels.

$$recon(t) = \sum_{i=1}^{M} recon_i(t) \tag{11}$$

III. Strategy Verification

In order to verify that the proposed stochastic impulse trains can convey the correct speech information, we used two Mandarin Chinese syllables /wan/ with tone (\) and /you/ with tone (\/). In this study, we used eight channels with central frequencies of 366Hz, 526Hz, 757Hz, 1089Hz, 1566Hz, 2252Hz, 3241Hz, and 4664Hz. All of the simulations were implemented in Matlab.

The syllable /wan/ (tone (\),which means "ten thousand," was processed by the zerocrossing-based strategy and the proposed strategy in order to demonstrate how the non-uniform impulse train works before the impulse train is amplified by the temporal envelope. We selected one frame to analyze the third channel. Figure 3(a) shows the waveform for one frame. Figure 3(b) shows the impulse train generated by the zerocrossing-based strategy for the same frame, and Fig.3(c) shows the impulse train generated by the proposed strategy for the same frame. Their corresponding spectra were also calculated. Figure 3(d) shows the spectrum of the original waveform, Fig. 3(e) shows the spectrum of the impulse train generated by the zerocrossing-based strategy, and Fig. 3(f) shows the spectrum of the impulse train generated by the proposed strategy. Non-uniform stimulation impulse trains were applied only to the LF channels, so we simply analyzed the spectrum of the LF range.

Based on a comparison of the impulse trains with the two strategies, it can be seen that using the zerocrossing-based strategy, the impulse generated at the zerocrossing time had little relationship with the waveform amplitude. By contrast with the proposed strategy, the impulse generation time was stochastic and it varied with the waveform amplitude. When the amplitude was lower, the impulses could be controlled by speech more effectively, where the amplitude was higher the impulses were more stochastic.

Based on a comparison of the spectra of the impulses generated by the two strategies and that of the original waveform, it can be seen that the proposed stochastic impulse train and the zerocrossing-based impulse train include some speech information. It is also clear that the proposed stochastic impulse train can convey the speech information more exactly. The amplitudes of the impulses are the same and the speech information is only conveyed by the interval times between the impulses; therefore, after modulation by the envelope, the speech information can be transmitted by the amplitude envelope but also by the interval times between the impulses in non-uniform stimulation strategies.

In addition, the syllable /you/ (tone (\/), which means "have," is utilized to demonstrate the functional significance of the proposed auditory model-based strategy and to illustrate how it works in Mandarin Chinese tones. Figure4 shows the pitch trajectories of the original syllable (a), the signal processed by CIS (b), the signal processed by the zerocrossing-based strategy (c), and processed by the proposed auditory model-based strategy (d).The program praat was used to extract the pitch trajectory. Clearly, the proposed strategy and the zerocrossing-based strategy extracted the pitch trajectory more exactly than the CIS strategy, implying that non-uniform stimulation strategies, including proposed auditory model based strategy and the zerocrossing-based strategy, can deliver more Mandarin tone information, thereby improving the perception of Mandarin.

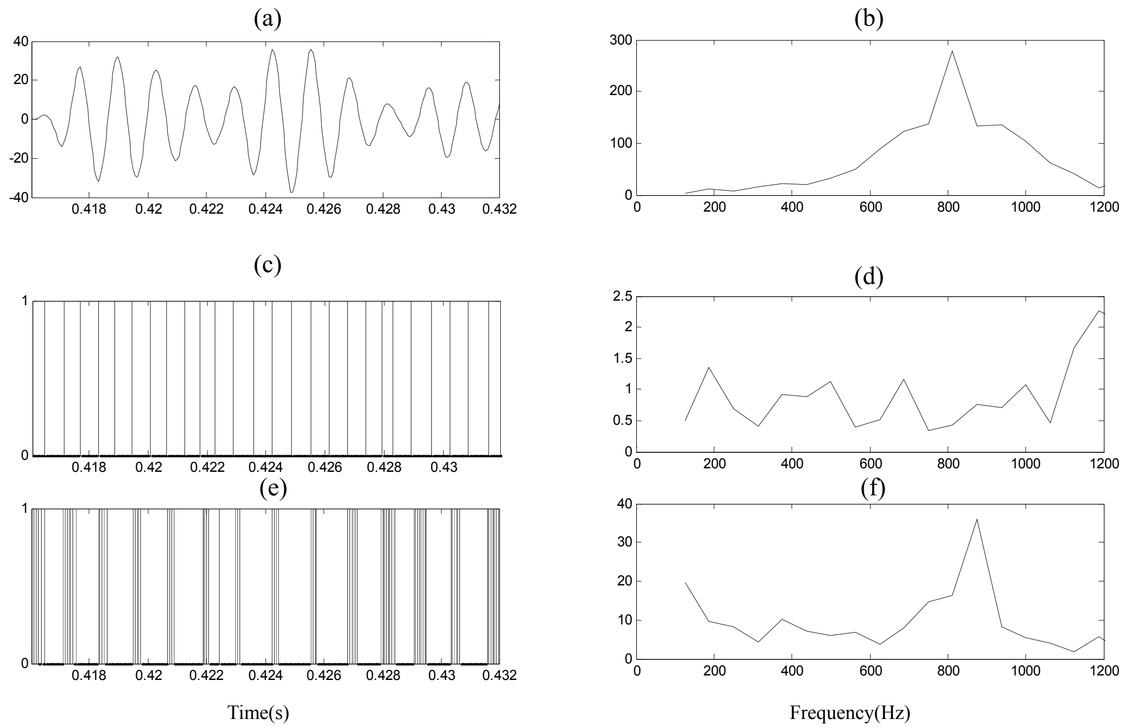

Fig.3. (a) Waveform of the selected frame and (d) its spectrum. (b) The impulse train generated by the zerocrossing-based strategy and (e)its spectrum. (c)The impulse train generated by the proposed strategy for the same frame and (f) its spectrum.

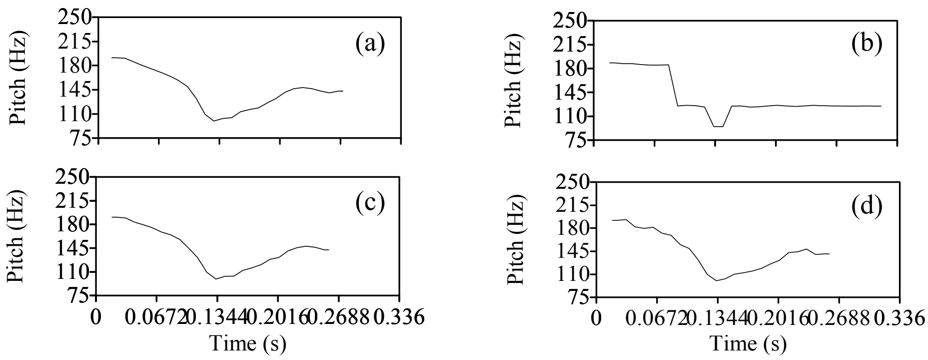

Fig.4. (a) Pitch trajectories of the original syllable and the signal processed by: (b) CIS, (c) the zero crossing-based strategy, and (d) the proposed auditory model-based strategy.

IV. Experimental Design

We performed three psychoacoustic experiments with normal-hearing listeners to evaluate the performance of the auditory model based on the stochastic impulse stimulation strategy for Mandarin identification in noisy and quiet environments. Experiment 1 examined Mandarin tone identification in quiet conditions and in the presence of steady-state white noise. Experiment 2 examined Mandarin word identification in quiet conditions and in the presence of steady-state white noise. Experiment 3 replicated more realistic listening conditions by measuring Mandarin sentence identification in quiet conditions and in the presence of white noise.

A. Subjects

In total, 18 normal-hearing, Mandarin-speaking subjects (ten male and eight female adults aged from 20–26 years) were recruited to participate in the experiments. The 18 subjects were divided into three groups each of six subjects. Each group participated in one of the three experiments. Each experiment tested four processing conditions (CIS, zerocrossing, proposed method, and natural speech) and four noise conditions (0dB, 5dB, 10dB, and clean).

B. Test Materials

The speech signals were selected from the listening ability test system database (Angel Sound TM). The Mandarin tone materials comprised six vowels (/a/, /o/, /e/, /i/, /v/, and /u/), each with four tones, which were spoken by a female adult. The Mandarin word materials comprised 100 words, including 25 consonant-vowel combinations, each with four tonal patterns. The selection of these words was based on the following two principles: maximizing the diversity of the consonant and vowel combinations, and ensuring that each consonant-vowel-tone combination was lexically meaningful in Mandarin Chinese [19]. The target sentence materials comprised 160 everyday expressions in Chinese, which were spoken by the same female adult. In the sentence experiment, there were eight randomized sentences for each condition.

C. Signal Processing

Eight channels with central frequencies of 366, 526, 757, 1089, 1566, 2252, 3241, and 4664 Hz were used in all three strategies (i.e., the CIS, zerocrossing-based, and proposed strategies). The low-pass filters used to extract amplitude envelope information in all experiments were two-order Butterworth filters with a cutoff frequency of 400 Hz. In the zerocrossing-based strategy and the proposed strategy, the three LF channels with central frequencies of 366, 526, and 757 Hz were selected to apply the corresponding non-uniform impulse stimulation. Thus, in the zerocrossing-based strategy, the impulse trains generated at the waveform zerocrossing time from each channel were used instead of the uniform impulse stimulation, and in the proposed strategy, the impulse trains generated at the auditory nerve firing time, which was stochastic, were used instead of the uniform impulse stimulation.

D. Procedure

Before the experiments, each subject attended a training session to make them familiar with the synthesized sounds. The experiments were conducted in a double-walled, sound-attenuated booth.

In the Mandarin tone identification experiment, after the presentation of the vowels with different tones, the subject was asked to complete a four-alternative forced-choice Mandarin tone identification task where they wrote down the number corresponding to the perceived tone. Finally, the percentage of correct scores was calculated for each condition. All of the stimuli were presented in a random order.

In the Mandarin word identification experiment, the subject listened to the processed sounds and wrote down the words they heard according to the 100-word table. Under each condition, 32 words were chosen randomly from the 100-word database. Finally, the percentage of correct scores was calculated for each condition.

In the Mandarin sentence recognition experiment, each subject was asked to write down as many words as possible from the target sentence after the sentence was presented. The number of correctly identified words was recorded to calculate the final percentage of correct scores for each condition.

V. Experimental Results

A. Mandarin Tone Identification

Figure 5 shows the Mandarin tone identification scores as a function of the signal-to-noise ratio (SNR) for four processing conditions and the *p*-values are shown in Table 1. In general, the zerocrossing-based strategy and the proposed strategy performed better than the CIS strategy. Analysis of variance showed that the performance was significantly better for the two non-uniform impulse stimulations, i.e., the zerocrossing-based condition ($p < 0.05$) and the proposed processing condition ($p < 0.05$), compared with the CIS strategy. The zerocrossing-based strategy and the proposed strategy perform almost as well as natural speech, and there was no significant difference among these three conditions ($p > 0.05$).

B. Mandarin Word Identification

Figure 6 shows the Mandarin word identification scores as a function of the SNR for the four processing conditions and the *p*-values are shown in Table 2. The results obtained in natural conditions are presented for reference. First, word identification with the natural stimulus was relatively resistant to noise over the SNR range that we tested, with almost perfect performance in clean conditions but it declined gradually to 89% at 0dB SNR. Second, the performance with the processed words was more than 50% lower than that with natural speech and it was sensitive to noise ($p < 0.05$). Third, the differences between each pair of the three strategies were significant (all $p < 0.05$). Therefore, the proposed strategy performs best compared with the other two processing strategies.

C. Mandarin Sentence Identification

Figure 7 shows the sentence identification scores as a function of the SNR from 0 to 10 dB as well as the clean conditions, and the *p*-values are shown in Table 3. The results obtained in natural conditions are also presented for comparison. It should be noted that sentence identification with the natural stimulus was almost 100% in any of the noise conditions considered in this study. The sentence identification scores were generally higher than the word identification scores with the three processing strategies, which was probably because the subjects had normal hearing and primary knowledge of the sentences. When the subjects heard some words of the sentences, they often knew the meanings of the sentences immediately. Therefore, similar to the word identification results, the proposed strategy performed significantly better than the zero crossing-based and CIS strategies ($p < 0.05$). Of the latter two strategies, the zero crossing-based strategy performed the best ($p < 0.05$). In addition, the SNR was a significant factor for all three strategies ($p < 0.05$). Finally, the average improvement was 27% for the proposed strategy compared with the CIS strategy, and 13% for the proposed strategy compared with the zero crossing-based strategy. These improvements in Mandarin sentence identification suggest that the information conveyed by the auditory model-based stochastic impulse intervals was important for improving the cochlear implant performance at Mandarin in realistic listening situations.

TABLE 1 Mandarin Tone Identification's P Values For Four Strategies

	natural	CIS	zerocrossing	proposed
natural	nil	< 0.05	< 0.05	0.20
CIS	< 0.05	nil	< 0.05	< 0.05
zerocrossing	< 0.05	< 0.05	nil	0.01
proposed	0.20	< 0.05	0.01	nil

TABLE 2 MANDARIN WORD IDENTIFICATION'S P VALUES FOR FOUR STRATEGIES

	natural	CIS	zerocrossing	proposed
natural	nil	< 0.05	< 0.05	<0.05
CIS	< 0.05	nil	< 0.05	< 0.05
zerocrossing	< 0.05	< 0.05	nil	<0.05
proposed	<0.05	< 0.05	<0.05	nil

TABLE 3 MANDARIN SENTENCE IDENTIFICATION'S P VALUES FOR FOUR STRATEGIES

	natural	CIS	zerocrossing	proposed
natural	nil	< 0.05	< 0.05	<0.05
CIS	< 0.05	nil	< 0.05	< 0.05
zerocrossing	< 0.05	< 0.05	nil	<0.05
proposed	<0.05	< 0.05	<0.05	nil

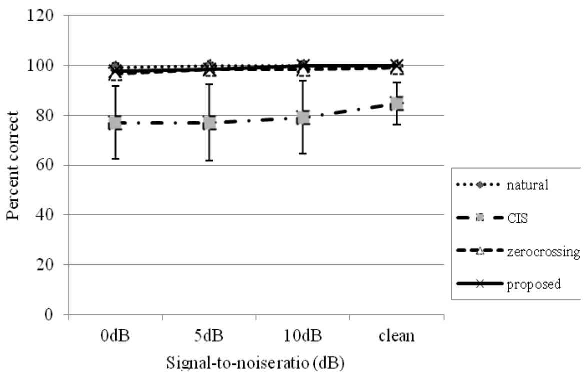

Fig.5. Mandarin tone identification scores as a function of SNR (x-axis) in the four processing conditions.

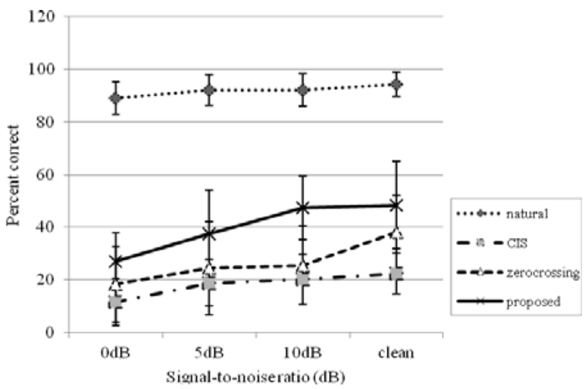

Fig.6. Mandarin word identification scores as a function of SNR (x-axis) for four processing conditions.

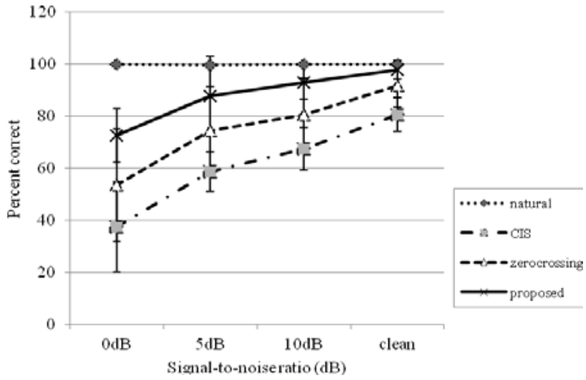

Fig.7. Mandarin sentence identification scores as a function of SNR (x-axis) in the four processing conditions.

VI. CONCLUSION

Only spatial coding is applied in modern strategies such as the CIS and NofM strategies. In our proposed stimulation strategy, the information transmitted to the central nervous system is coded in the envelope, but also in the intervals between the stimulation impulses in the same manner as the natural system. Cochlear stimulation using the proposed method has the advantage of simulating the auditory nerve behavior in a more natural manner.

Mandarin tone identification by normal-hearing listeners demonstrated that speech reconstruction by the proposed strategy and the zerocrossing-based strategy obtained better tone recognition scores than that processed using uniform stimulation, such as the CIS strategy, and these methods performed almost as well as natural speech, even in noisy conditions. The proposed strategy achieved the best Mandarin word and sentence identification scores in any SNR conditions, thereby suggesting that the proposed strategy can deliver the most Mandarin information to the auditory system, and thus improve the perception of Mandarin Chinese.

Overall, our results indicate that the proposed strategy based on a consideration of the peripheral auditory system is a powerful technique for Mandarin Chinese recognition and perception.

REFERENCES

[1] Lan, N., Nie, K.B., Gao, S.K. Zeng, F.G., 2004. "A novel speech-processing strategy incorporating tonal information for cochlear implants," IEEE Transactions on Biomedical Engineering 51(5), 752-760.

[2] Smith, Z.M., Delgutte, B., Oxenham, A.J., 2002. "Chimaeric sounds reveal dichotomies in auditory perception," Nature 416, 87-90.

[3] Eddington, D.K., Dobelle, W.H., Brackmann, D.E., Mladevosky, M.G., Parkin, J.L., 1978."Auditory prosthesis research with multiple channel intracochlear stimulation in man," The Annals of Otology, Rhinology,

and Laryngology 87, 1-39.

[4] Wilson, B.S., Finley, C.C., Lawson, D.T., Wolford, R.D., Eddington, D.K., Rabinowitz, W. M., 1991. “Better speech recognition with cochlear implants,” Nature 352, 236-238.

[5] McDermott, H.J., McKay, C.M., Vardal, A.E., 1992. “A new portable sound processor for the University of Melbourne/Nucleus limited multi-electrode cochlear implant,” Journal of the Acoustical Society of America 91, 3367-3371.

[6] Nie, K.B., Stickney, G., Zeng, F.G., 2005. “Encoding frequency modulation to improve cochlear implant performance in noise,” IEEE Transactions on Biomedical Engineering 52, 64-73.

[7] Zeng, F.G., 2002, “Temporal pitch in electric hearing,” Hearing Research 174, 101-106.

[8] Matousek, J., Tihelka, D., Romportl, J., Psutka, J. Slovak “unit-selection speech synthesis: Creating a New Slovak voice within a Czech TTS system ARTIC,” IAENG International Journal of Computer Science, vol. 39, no. 2, pp147-154 , 2012

[9] Fearn,R., Wolfe,J., 2000. “Relative importance of rate and place: experiments using pitch scaling techniques with cochlear implant recipients,” Annals of Otology Rhinology & Laryngology Supplement 185(185),51–53.

[10] Zeng,F.G., Nie,K., Stickney,G.S., Kong,Y.Y., Vongphoe,M., Bhargave, A., Wei,C., Cao,K., 2005. “Speech recognition with amplitude and frequency modulations,” Proceedings of the National Academy of Sciences 102 (7), 2293–2298.

[11] Chen, F., Zhang, Y.T., 2011. “Zerocrossing-based nonuniform sampling to deliver low-frequency fine structure cue for cochlear implant,” Digital Signal Processing 21, 427-432.

[12] Luo, X., Fu, Q.J., 2006. “Contribution of low-frequency acoustic information to Chinese speech recognition in cochlear implant simulations,” Journal of the Acoustical Society of America 120, 2260-2266.

[13] Wilson, B.S., Lawson, D.T., Muller, J.M., Tyler, R.S., Kiefer, J., 2003. “Cochlear implants: Some likely next steps,” Annual Review of Biomedical Engineering 5, 207-249.

[14] Martinez-Rams, E.A., Garceran-Hernandez, V., 2007. “ANF stochastic low rate stimulation,” Lecture Notes in Computer Science 4527, 103-112.

[15] Li, Z.H., Chi, H.S., 2006. “Progress in computational modeling of auditory periphery,” Acta Acustica 31(5), 449-465.

[16] Martinez-Rams, E.A., Garceran-Hernandez, V., Ferrandez-Vicente, J.M., 2009. “Low rate stochastic strategy for cochlear implants,” Neurocomputing 72, 936-943.

[17] Samborski, R., Ziolko, M., ”Filter-based model of multimicrophone array in an adverse acoustic environment, ” Engineering Letters, vol. 20, no. 4, pp336-338, 2012

[18] Chen, F., Zhang, Y.T., 2008. “Zero-crossing-based fine structure representation to convey Mandarin tonal information: A study on the noise effect,” Engineering in Medicine and Biology Society (EMBS), 30th Annual International Conference of the IEEE. 343-346.

[19] Wei, C.G., Cao, K.L., Zeng, F.G., 2004. “Mandarin tone recognition in cochlear-implant subjects,” Hearing Research 197. 87-95.