A Novel Speech-Processing Strategy Incorporating Tonal Information for Cochlear Implants

N. Lan*, K. B. Nie, S. K. Gao, and F. G. Zeng

***Abstract*—Good performance in cochlear implant users depends in large part on the ability of a speech processor to effectively decompose speech signals into multiple channels of narrow-band electrical pulses for stimulation of the auditory nerve. Speech processors that extract only envelopes of the narrow-band signals (e.g., the continuous interleaved sampling (CIS) processor) may not provide sufficient information to encode the tonal cues in languages such as Chinese. To improve the performance in cochlear implant users who speak tonal language, we proposed and developed a novel speech-processing strategy, which extracted both the envelopes of the narrow-band signals and the fundamental frequency (F_0) of the speech signal, and used them to modulate both the amplitude and the frequency of the electrical pulses delivered to stimulation electrodes. We developed an algorithm to extract the fundatmental frequnency and identified the general patterns of pitch variations of four typical tones in Chinese speech. The effectiveness of the extraction algorithm was verified with an artificial neural network that recognized the tonal patterns from the extracted F_0 information. We then compared the novel strategy with the envelope-extraction CIS strategy in human subjects with normal hearing. The novel strategy produced significant improvement in perception of Chinese tones, phrases, and sentences. This novel processor with dynamic modulation of both frequency and amplitude is encouraging for the design of a cochlear implant device for sensorineurally deaf patients who speak tonal languages.**

***Index Terms*—Acoustic signal processing, cochlear implants, electrical stimulation, speech perception, tonal language.**

I. Introduction

COCHLEAR implant (CI) devices have been successful in restoring hearing to profoundly deaf patients through electrical stimulation of the auditory nerve with fine electrodes inserted into the scala tympani of the cochlea [19], [21]. The performance of cochlear implants depends in large part on the speech processor to faithfully decompose speech signals into a number of channels of narrow-band electrical signals that may be used to activate the spiral ganglion cells of the auditory nerve. The number of electrodes in modern cochlear implant devices may be different from the number of channels in the speech processor, and may vary from 12 to 22 with monopolar or bipolar arrangements. The configuration and placement of the electrodes are of importance to the overall performance of these devices. Recently, understanding on what speech features should be extracted and how to encode them for stimulation has significantly improved the speech recognition for cochlear implant users [20].

In general, the processing strategies of speech signals may extract and encode temporal and/or spectral cues in the speech for cochlear implant devices [1], [22], [23]. In the early Nucleus device, the fundamental frequency (F_0) was used to modulate the pulse rate proportionally during voiced sound, and spectral information with one (F2) or two formants (F1, F2) from the speech signal was used to select stimulation electrodes, respectively [1], [22], [23], [30]. Such strategies were able to achieve an open-set of speech recognition, a significant improvement over the single electrode devices. The compressed analog (CA) algorithm decomposed the speech signal into a few narrow-band signals for simultaneous stimulation of the auditory nerve with multiple electrodes [12]. Theoretically, the CA strategy used both temporal and spectral information in the original speech signal. However, simultaneous stimulation across electrodes often resulted in electric field interaction that was detrimental to speech recognition [1], [22], [23], [36]. To overcome this problem, the continuous interleaved sampling (CIS) strategy [36] was proposed to avoid channel interactions during stimulation, in which the envelope cues of the bandpass filtered speech signals were extracted and encoded to modulate the amplitude of stimulation pulses. The CIS strategy showed a high level of speech recognition for the cochlear implant users of monotonal languages, such as English and German [3], [29], [35]–[37].

While the present cochlear implant devices have achieved a high rate of perception for English speaking users, it was reported that cochlear implant users who spoke Chinese showed poor results in identification of vowels and consonants compared with English speaking users [44], [45]. This is because Chinese is a tonal language that uses four basic tones [1] to express different meanings of words. Experimental evidences also indicated that several significant differences existed between Chinese and English speech recognition with CIS strategy [15]. Recently, Xu *et al.* [40] identified that the spectral details of the filtered signals played a more important role in speech perception for Chinese. Their experimental results with native Chinese speaking subjects revealed that the tonal information in Chinese speech was encoded primarily in fine details in the spec-

Manuscript received March 11, 2002; revised July 25, 2003. The work of N. Lan was supported by a Grant from Li Foundation of San Francisco. The work of K. B. Nie was suported in part by the Natural Science Foundation of China (CNSF) under Grant 30000041. The work of F. G. Zeng was supported in part by the National Institutes of Health (NIH) under Grant 2RO1DC02267. *Asterisk indicates corresponding author.*

*N. Lan is with the Department of Biokinesiology and Physical Therapy, University of Soutern California, CHP-155, Los Angeles, CA 90089 USA (e-mail: ninglan@usc.edu).

K. B. Nie and F. G. Zeng are with the Departments of Otolaryngology and Biomedical Engineering, University of California, Irvine, CA 92697 USA.

S. K. Gao is with the Department of Biomedical Engineering, Tsinghua University, Beijing 100084, China.

Digital Object Identifier 10.1109/TBME.2004.826597

[1]Some dialects in China, e.g., Cantonese, use as many as six tones.

0018-9294/04$20.00 © 2004 IEEE

trum of the speech signal. Thus, the current speech processors of cochlear implant devices may not meet the needs of a large population of users in countries of tonal languages.

The objective of this work is to test the hypothesis that modulating the pulse rate of stimulation according to the pitch variation of tones in combination with an envelope extraction processor, such as CIS, can significantly improve the intelligibility of Chinese speech by cochlear implant users. In Section II, we present the spectral characteristics of pitch variations in Chinese speech, and describe a design of the novel speech processor that includes extraction of tonal information and dynamical coding of stimulation frequency. Spectrum analysis and computer simulation of the CIS and novel strategies are elucidated in Section III. Experiments in native Chinese speaking subjects with normal hearing are described in Section IV. Results obtained in this study are explained in Section V. The relevance and implication of these results in the design of an effective cochlear implant device that provides high discrimination of tonal cues are discussed in Section VI. Preliminary results are also reported elsewhere [26], [27].

II. A Novel Speech Processor

A. Patterns of Tonal Variation

Chinese is a tonal language, whose semantics depends on particular patterns of pitch variation in the pronunciation of words. For example, a single syllable word "ma" can be pronounced with four tonal patterns, a flat tone $(-)$, a rising tone $(/)$, a falling–rising tone $(\vee)$, and a falling tone $(\backslash)$. Each tone yields a different meaning, e.g., the flat tone could mean "mother," the rising tone "numbness," the falling– rising tone "horse," and the falling tone "scolding," respectively. The fundamental frequency (i.e., F_0) of the four different tones of "ma" is presented in Fig. 1. It clearly shows the distinct patterns in the fundamental frequency F_0 with the four different tones. Generally, the flat tone is associated with a constant frequency, the rising tone with a rising frequency, the falling-rising tone with a falling-rising frequency and the falling tone with a falling frequency. These attributes in F_0 may be important in Chinese perception by cochlear implant users.

Another notable characteristic of Chinese speech is its regularity in phonetics. Most characters (or words) in Chinese are a single syllable unit composed of two phonemes, a consonant followed by a long vowel. In pronunciation, the consonant introduces the word in a short period of time, and the vowel is sustained in a relatively long period of time, during which the four different tones are produced. A phrase is formed with two or more words, and a sentence is a sequence of words or phrases grouped in a grammatical structure. This salient feature of phonetics makes it possible to extract tonal patterns of words, and use them to modulate the stimulation frequency of cochlear implants. In this study, we developed a novel processing strategy that modulated the center frequency of stimulation of a CIS processor according to the F_0 pattern of tones of Chinese words.

B. Novel Speech Processor

Fig. 2 illustrates the novel speech processor that uses both CIS of bandpass filtered signals and pulse-rate modulation

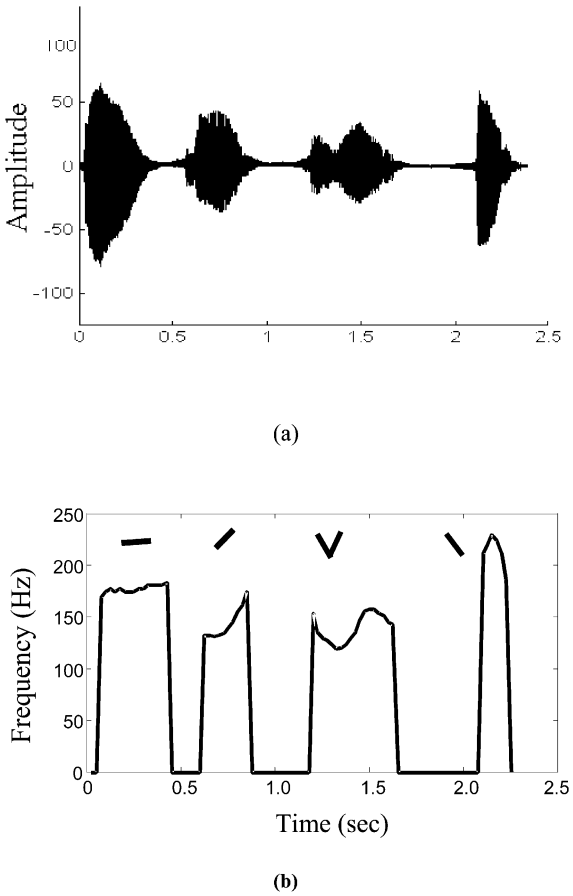

Fig. 1. Tonal patterns in Chinese speech. (a) Original speech waveform of a Chinese word with four different tones, each of which has different meanings, from left to right, "mother" $(-)$, "numbness" $(/)$, "horse" $(\vee)$, and "scolding" $(\backslash)$, respectively. (b) Corresponding fundamental frequency (F_0) as a function of time. F_0 is extracted from the original speech signal using the algorithm described in Section II-C.

based on pitch variation of tones. The novel speech-processing strategy has two signal pathways, including the traditional envelope extraction and an additional fundamental frequency processing (Fig. 2). The envelope-extraction method is similar to the standard CIS strategy. A sound signal is sampled, pre-emphasized, and then decomposed into multiple frequency bands by a bank of bandpass filters. The filtered signal is rectified and then smoothed by a low-pass filter to extract the envelope. In cochlear implants, the band-specific envelope is also logarithmically compressed and then used to modulate the amplitude of biphasic pulse trains that are interleaved among electrodes [36]. The interleaved sampling strategy is effective to avoid stimulation interference among channels, and thus enhancing the performance of speech perception significantly [31], [32], [35].

The second pathway in the novel strategy explicitly extracted F_0 by a specially designed algorithm described in Section II-C (Fig. 3). An artificial neural network was used to verify the extracted tonal patterns (Section II-D). The dynamic coding of pulse rate was explained in acoustic simulation (Section III), where F_0 was used to modulate the center frequency of sinusoidal waves. In actual implementation, the F_0 will be used to modulate the rate of stimulation pulses around a center frequency in a similar way as described in (2) of Section III-A.

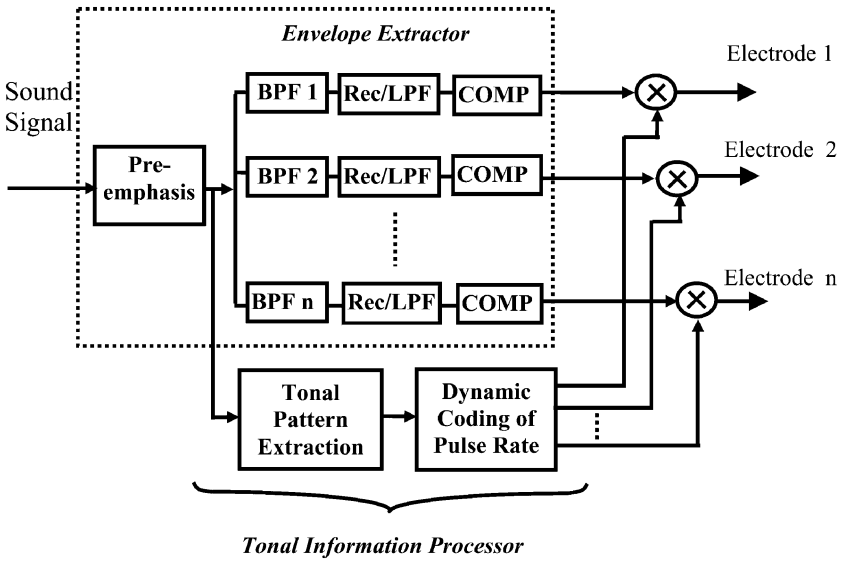

Fig. 2. Block diagram of the novel strategy incorporating tonal infomration. Inside the dotted line box is the CIS strategy that includes a pre-emphasis filter, bandpassed filters (BPF), full-wave rectifiers (Rec), low-passed filters (LPF) for the envelope extraction, and an amplitude compressor (COMP). A separate tonal information processor is used to extract tonal patterns from the pre-emphasized speech signal, and dynamically code the stimulation rate of pulse train. The amplitudes of the pulse trains carrying tonal information in their pulse rate are modulated by the output envelope of CIS processor, and are delivered to each electrode. This makes it possible to transmit tonal information (F_0) to auditory nerves using synchronized pulse rate.

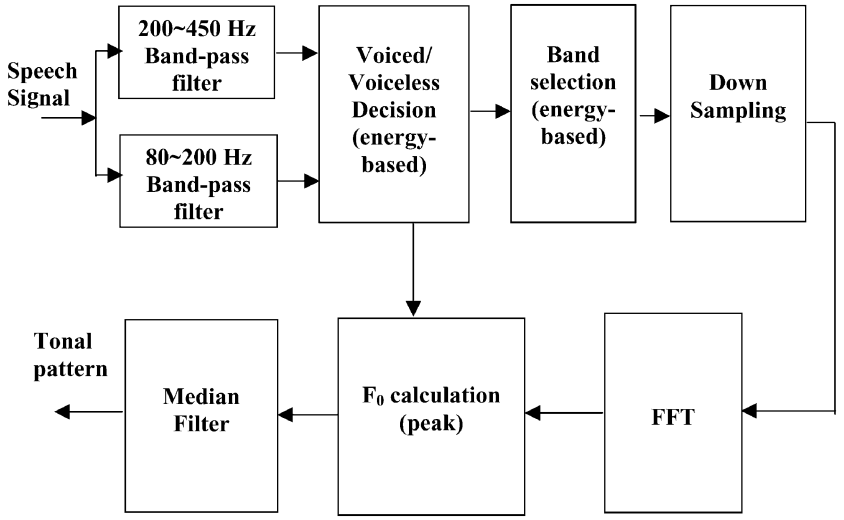

Fig. 3. Algorithm for fundamental frequency (F_0) extraction. A speech signal is divided into two subbands, 80–200 Hz and 200–450 Hz. The signal is then classified as either a voiced or voiceless sound. In the case of the voiceless sound, F_0 is set to 0 Hz. In the case of the voiced sound, the subband signal with the highest energy is selected, down-sampled, and analyzed in frequency domain by FFT to calculate F_0. The length of data window for spectral analysis is 512 points (or 46.4 ms). Moving windows with 256 points of overlap are used in the sequential analysis of F_0 in time. The F_0 trajectory is smoothed with a 3-point median filter.

C. Extraction of Pitch Information

In Chinese speech, the F_0 of a male voice is mainly confined within 80 to 200 Hz, while that of the female voice lies between 200 and 450 Hz [42]. Thus, we used a two-channel fast Fourier transform (FFT)-based F_0 extraction algorithm that separately processed male and female voices. In the proposed algorithm, two bandpass filters first divided the male and female voices into two separate channels. In each channel, the root mean square (rms) levels combined with zero crossing detection were used to determine whether a segment of the signal was voiced or voiceless. For the voiceless sound, F_0 was set to zero. For the voiced sound, the channel with a higher energy level was selected and then down-sampled from 11 025 to 1378 Hz. The voiced signal sampled at 11 025 Hz was segmented with an analysis window of 512 points of duration (about 46.4 ms). An FFT algorithm in MATLAB was used to compute the spectrum of the segmented signal, and the F_0 was obtained from the FFT spectrum's peak location. A moving window with 256 data point overlap was used to generate a time series of F_0 trajectory. Thus, the inter-frame duration between windows was actually of 23.2 ms. In addition, a 3-point median filter was used to smooth the F_0 trajectory during the voiced sound. Fig. 5 shows an example of the extracted F_0 trajectory from a sentence.

D. Varification of Extracted Tonal Patterns

The extraction of the four tones from the F_0 patterns is varified with a three-layered feedforward artificial neural network

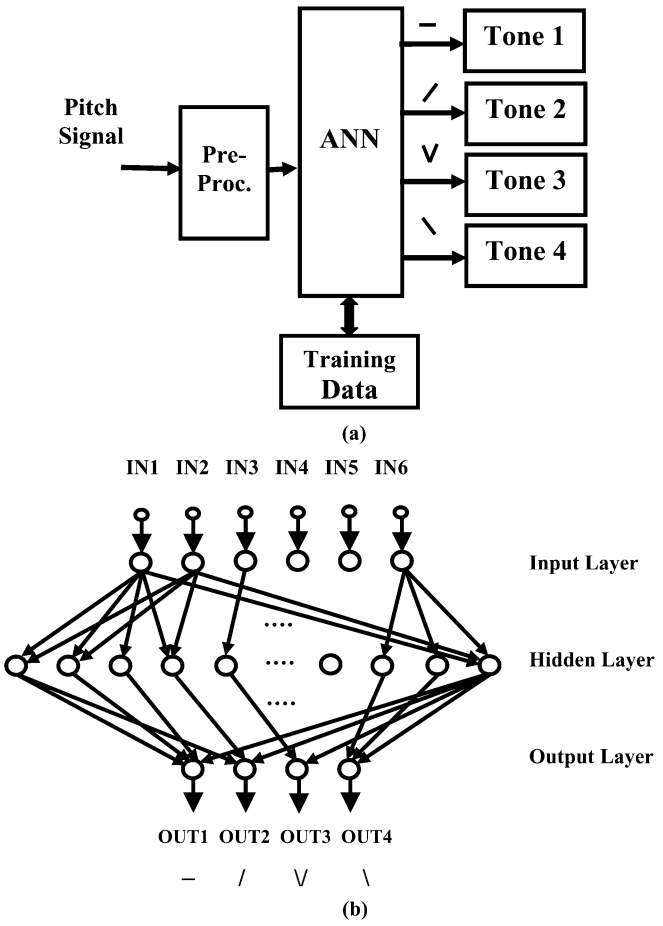

Fig. 4. Automatic identification of tonal patterns. (a) Block diagram of an automatic classifier of tonal patterns, in which an ANN is the core for tonal pattern differentiation. The identified tonal pattern could then be used for coding the frequency of carrier pulses. (b) A 3-layer forward ANN is used in this implementation. The input layer has six inputs accepting equaly spaced six points from the tonal trajectory. The hidden layer has 12 neurons, and the output layer has four neurons corresponding to the four tonal patterns. The activation function of these neurons is the nonlinear, sigmoidal activation functions. At the output layer, the tone is determined by the output with the largest value.

(ANN) trained to recognize the four tonal patterns. The ANN has a three-layered feedforward structure (Fig. 4), in which there are six neurons in the input layer, 12 neurons are embedded in the hidden layer, and four output neurons correspond to the four tones in Chinese speech. The values of six frames of F_0 evenly distributed within the tonal trajectory of the word were used as the input to the neural network. A back-propagation algorithm was used to train the ANN using sample words of various tonal variations. In the idenfication of tones, the largest output of ANN corresponded to one tonal pattern out of the four possible tones. For example, an output vector of $[1\ 0\ 0\ 0]$ would suggest a flat tone $(-)$. In this study, the ANN was not implemented in real time, and the effects of the delay were not evaluated.

III. Acoustic Simulation

A. Acoustic Simulation

Computer simulation of the CIS and the novel processors was carried out using MATLAB. For both the CIS and the novel processors, a pre-emphasis filter (1.2-kHz high pass) was used to provide spectral equalization above 1.2 kHz. The dynamic range compressors in the CIS processor and the novel processor were disabled in simulations. The bandpass filters implemented in both processors were six-order Butterworth filters, and the envelope was extracted with full-wave rectification followed by a fourth-order Butterworth type low-pass filter. The cutoff frequency in the envelope filter was 50 Hz. Pre-emphasis and de-emphasis filters were first-order FIR filters.

The processed signals were re-synthesized into voice sound, and presented to subjects with normal hearing in experimental tests [2], [6], [8]–[10]. Synthesis of voice was achieved using superimposed sinusoidal signals in each channel [21]

$$s(t) = \sum_{k=1}^{N} A_k(t) \sin(2\pi f_k(t) + \Phi) \tag{1}$$

in which $s(t)$ is the synthesized sound signal, subscript k indicates the channel number, N is the total number of bands, $A_k(t)$ represents the envelope of the kth band, $f_k(t)$ is a time-varying function of stimulation frequency, and Φ is the initial phase estimated from FFT. The updating rate for A_k and f_k in simulation is in accordance to the inter-window time frame of 23.2 ms.

In the acoustic simulation of the CIS strategy, the envelope signal $A_k(t)$ was modulated by a pure sinusoidal signal, and its center frequency $f_k(t)$ was set at a constant value (f_C) in each channel [8], [13]. Alternatively, $A_k(t)$ can be modulated by band-limited noise [8], [29]. To simulate the novel strategy, the center frequency $f_k(t)$ was dynamically varied as a function of the extracted F_0

$$f_k(t) = f_{Ck} + (F_0 - 200) \tag{2}$$

where f_{Ck} is the center frequency of each band, and F_0 is obtained through pitch extraction algorithm. Since the range of F_0 was from 80 to 450 Hz, F_0 was biased by 200 Hz to give a dynamic range of frequency modulation around f_{Ck}. The center frequency f_{Ck} of each band is chosen as the median frequency of the corresponding band (see Section IV-A).

B. Spectrum Comparison

To shed light on the ability of speech signal processing strategies to encode tonal information, preliminary analysis was performed to compare the spectrum of the original signal with those of the re-synthesized signals processed by either the CIS or the novel strategy. Spectral comparison was performed using the original signal of a Chinese sentence of a female voice, and the synthesized signals processed by a 4-channel CIS processor and a 4-channel novel processor. This should provide additional evidence to compare the novel strategy with the CIS strategy in encoding the tonal information embedded in Chinese speech.

IV. Experiment Tests

A. Test Materials

Experiments were conducted to compare the performance of the novel speech-processing strategy with that of the CIS in 20 subjects with normal hearing. Ten different single tones, 20 phrases, and 30 sentences were recorded in quiet environment pronounced with both male and female voices. The single tones included a variety of combination of consonants and vowels.

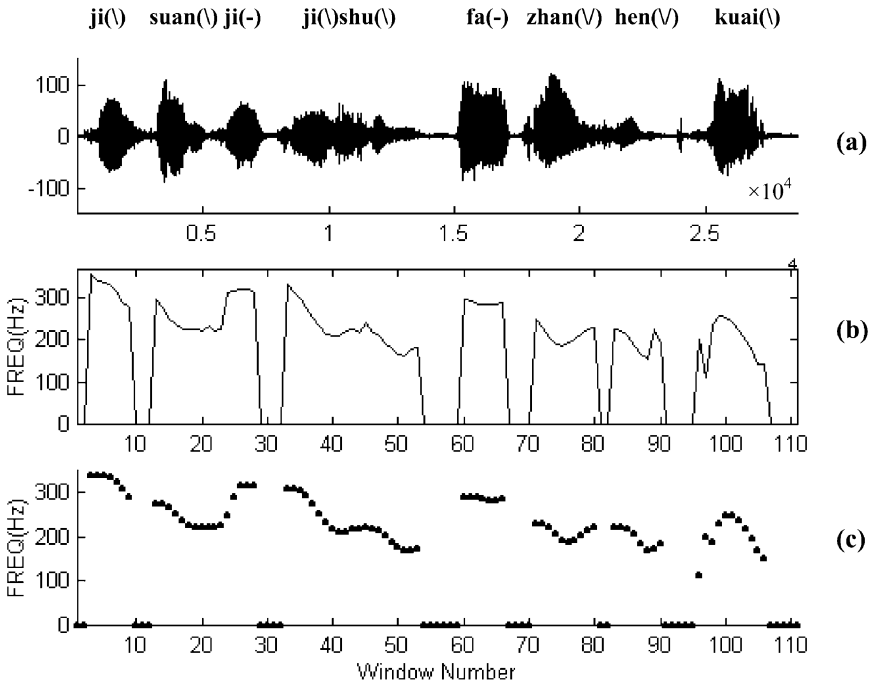

Fig. 5. Tonal trajectory of a sentence of Chinese speech. (a) Speech waveform of Chinese sentence, "ji (\) suan (\) ji (−) ji (\) shu (\) fa (−) zhan (\/) hen (\/) kuai (\)," which was recorded with a female speaker and sampled in 11.025 kHz. (b) Its raw tonal trajectory as a function of analysis window number. The analysis window length is of 512 points with 256-point overlap. (c) Smoothed trajectory processed with a medium filter.

The 20 phrases were chosen from most frequently used words in daily life. The 30 sentences were selected with no preference to a particular tone, and were used to evaluate both processors.

The voice signals for the test were sampled at 11 025 (Hz) and processed with 4, 6, and 8 channels, respectively, in the two strategies. The cutoff frequencies of bandpass filters in the 4-channel processor were 300, 620, 1285, 2657, and 5500 (Hz). They were 300, 486, 791, 1284, 2086, 3386, and 5500 (Hz) in the 6-channel strategy, and 300, 460, 800, 1385, 2057, 2800, 3900, 4500, and 5500 (Hz) in the 8-channel strategy. The processed signals were resynthesized by the computer (Section III-A) and the computer generated voice was presented to subjects during recognition tests.

B. Experiment Protocol

12 males and 8 females adult subjects (from 20 to 40 years old) with normal hearing participated in these experiments. Before the test, a 5-min training session was administered to all subjects. In the test for tone recognition, the subjects were given single tones in random orders. Computer generated voice of each tone was played to the subject once at a comfortable loudness. The subject was to select one correct answer from a set of multiple choices of four entries. In the test with sentences, a full sentence was played to the subject, who was then asked to write down the sentence just heard. For each subject, two sets of tests were administered, one with CIS strategy and one with the novel strategy. With each strategy, speech processors with 4, 6, and 8 channels were used in processing the sound signals.

C. Statistical Analysis

The paired t-test was used to detect the difference in the performance between the CIS and the novel strategies. A scale of 100% was used to quantify the performance of the subjects in identifying the tones, phrases or sentences. The test results from the CIS strategy and the novel strategy were paired, and the mean value of the difference of the correct rates between the two groups was calculated in the t-test. The null hypothesis stated that there was no difference in the performance between the two groups. The null hypothisis was accepted or rejected at a significance level of 95% confidence ($\alpha = 0.05$). Test results were summarized in Table II.

TABLE I
AUTOMATIC IDENTIFICATION OF TONAL PATTERNS

TONE		OUT1 (−)	OUT2 (/)	OUT3 (\/)	OUT4 (\)
计	(\)	0.3555	0.0000	0.2122	**0.4164**
算	(\)	0.0181	0.0000	0.0178	**0.9527**
机	(−)	**0.8927**	0.0885	0.0000	0.0050
技	(\)	0.0154	0.0000	0.0232	**0.9458**
术	(\)	0.0217	0.0000	0.0143	**0.9628**
发	(−)	**0.9732**	0.0001	0.0025	0.0672
展	(\/)	0.0000	0.0045	**0.9884**	0.0061
很	(\/)	0.0050	0.0000	**0.7022**	0.1054
快	(\)	0.0081	0.0000	0.0080	**0.9983**

V. RESULTS

A. Varification of Tonal Patterns

Table I demonstrates that the F_0 extraction algorithm provided sufficient information for the ANN to accurately identify

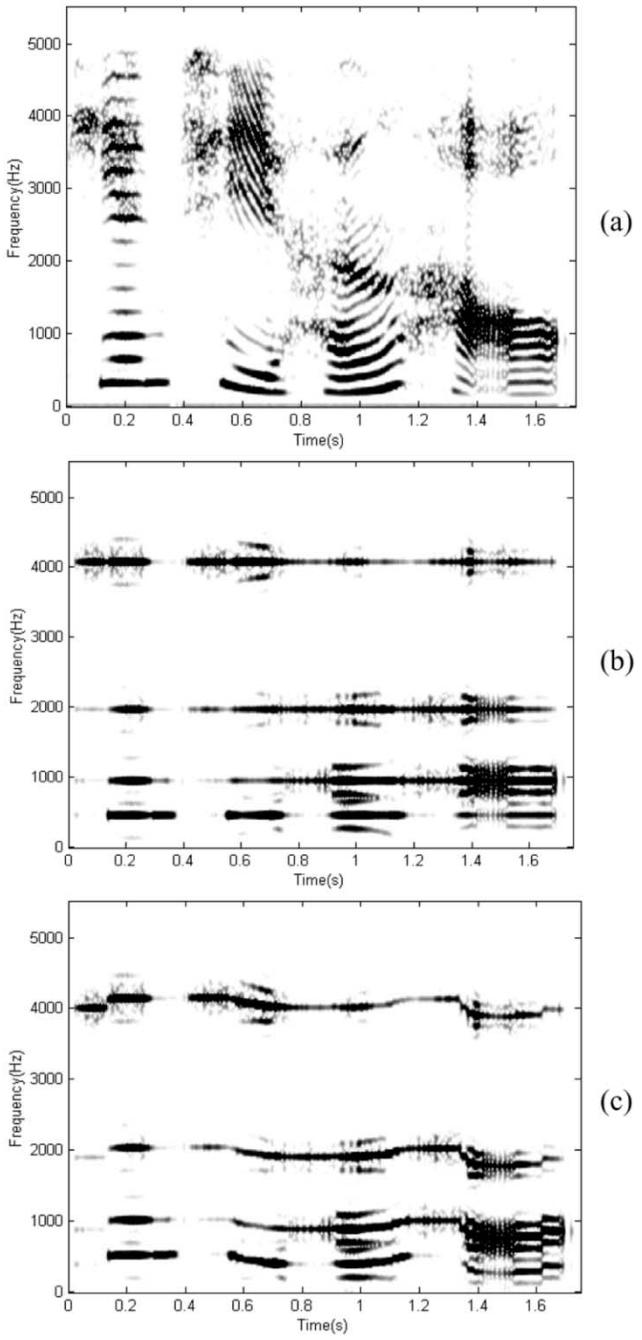

Fig. 6. Spectral comparison of the original and synthesized speech sounds. (a) Spectrum of the original signal. (b) Spectrum of the synthesized signal by a 4-channel CIS processor. (c) Spectrum of the synthesized signal by a 4-channel novel processor, respectively. The speech signal was from a female voice, speaking: "tian (−) qi (\) hen (\/) hao (\/)."

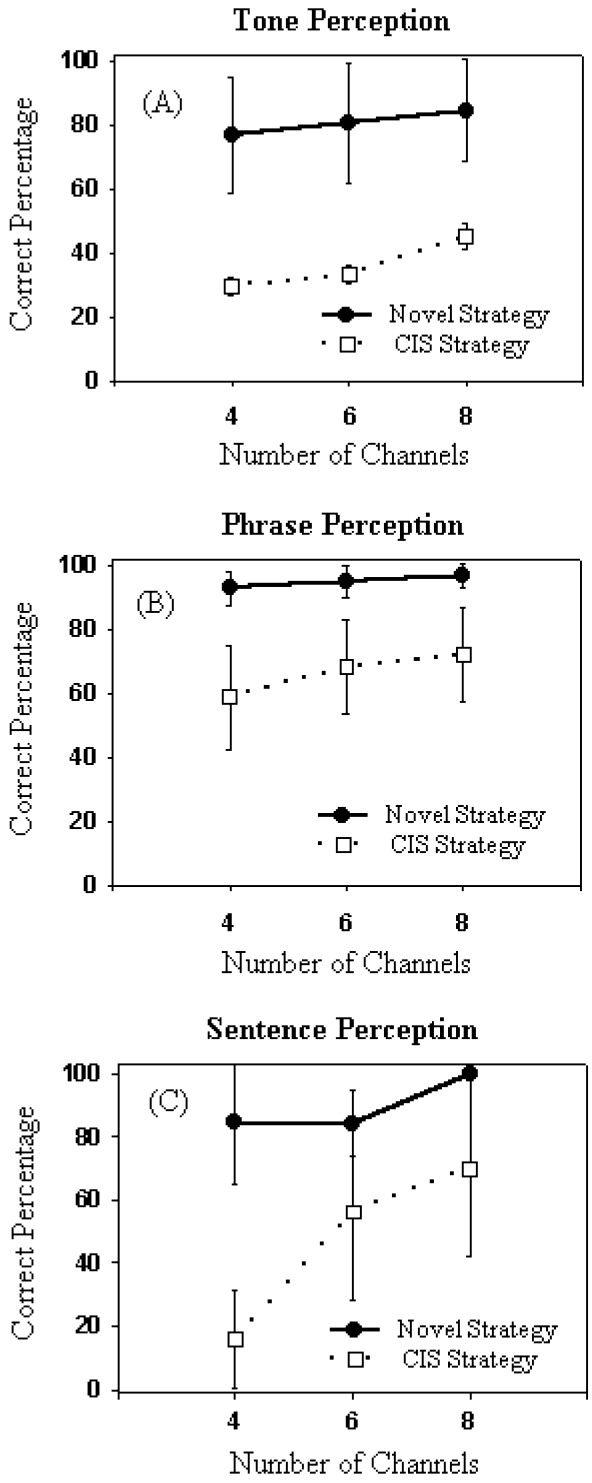

Fig. 7. Test results for the CIS strategy (open squares) and the novel strategy (filled circles) with 20 subjects with normal hearing function. The error bars represent the standard deviations in the tests. (a), (b), and (c) give the percentage of correct scores in identifying the 10 tones, 20 phrases, and 30 sentences with 4, 6, and 8 channels in each processor, respectively.

the tonal patterns. In this case, a sentence of nine words was examined, whose trajectories of fundamental frequencies (Fig. 5) were digitized, and used by the ANN for tonal pattern identification. The tonal classification was based on the winner-take-all decision rule with the largest output (underlined and bold) determining the tone. The Chinese characters are shown on the left side in Table I and their corresponding tones are given inside the parentheses. The ANN's outputs in response to these characters are tabulated in the right. In this test, the ANN was able to use the extracted F_0 to correctly discriminate tonal patterns.

B. Spectrum Comparison

Fig. 6 shows the spectrums of an original Chinese sentence and those processed by both the CIS and novel processing strategies with 4 channels. In the original speech spectrum of Fig. 6(a), frequency changes were clearly identified near

TABLE II
RESULTS OF THE PAIRED, TWO-TAILED t-TEST

No. of Ch. / Tests	4 Channels	6 Channels	8 Channels
Tones	$p = 1.1*10^{-7}$ (<0.01) H(-)	$p = 7.7*10^{-9}$ (<0.01) H(-)	$p = 1.72*10^{-7}$ (<0.01) H(-)
Phrases	$p = 4.4*10^{-9}$ (<0.01) H(-)	$p = 3.3*10^{-8}$ (<0.01) H(-)	$p = 1.5*10^{-7}$ (<0.01) H(-)
Sentences	$p = 1.1*10^{-10}$ (<0.01) H(-)	$p = 2.8*10^{-5}$ (<0.01) H(-)	$p = 1.3*10^{-4}$ (<0.01) H(-)

H: null hypothesis; (-) rejected; (+) accepted.

0.6 s (downward sweep) and 1.0 s (upward sweep). In the CIS-processed speech spectrum of Fig. 6(b), no apparent frequency changes were identified, and the loss of spectrum details was significant. In the spectrum processed by the novel strategy in Fig. 6(c), similar frequency change patterns were preserved although some spectral details were also lost. This indicates that the 4-channel novel processor is indeed superior to the CIS processor in encoding spectrum details of Chinese speech signals.

C. Experimental Tests

Fig. 7 shows the comparison of performance by the two strategies with 4, 6, and 8 channels for tone, phrase, and sentence perceptions. In tone perception, the novel strategy performed consistently better than the CIS strategy [Fig. 7(a)]. The correct rate for the novel strategy is about 80%, while the CIS was about 30%, indicating that tonal information in the novel strategy provided important cues for the subjects. For phrase recognition [Fig. 7(b)], the novel strategy achieved a higher correct rate of perception of about 90% as compared to a 60%–70% of success rate with the CIS. In sentence perception [Fig. 7(c)], the novel strategy demonstrated again a higher correct rate than the CIS. The paired t-test results in Table II confirmed that the difference in the performance between the two strategies was significant in all tests for tone, phrase, and sentence perception.

Fig. 8 presents the comparison of the effects of the number of channels in the two strategies on Chinese speech perception. With 4 channels, shown Fig. 8(a), the novel strategy demonstrated a much higher correct rate than the CIS strategy. In particularly, the 4-channel CIS processor showed a sharply lower correct rate (less than 20%) for sentence perception compared to the better-than-80% correct rate by the 4-channel novel strategy. With 6 and 8 channels, shown in Fig. 8(b) and (c), the success rate of the novel strategy for tones, phrases and sentences all show significant improvement than that of the CIS strategy. Note that the 8-channel novel processor Fig. 8(c) achieved a near 100% correct rate for both phrase and sentence perceptions.

VI. DISCUSSIONS

The test results in human subjects with normal hearing demonstrate that the novel strategy is consistantly superior than the CIS strategy for tone, phrase and sentence perception in Chinese speech. One reason for this limitation of the CIS strategy may be that it encodes only the amplitude envelope

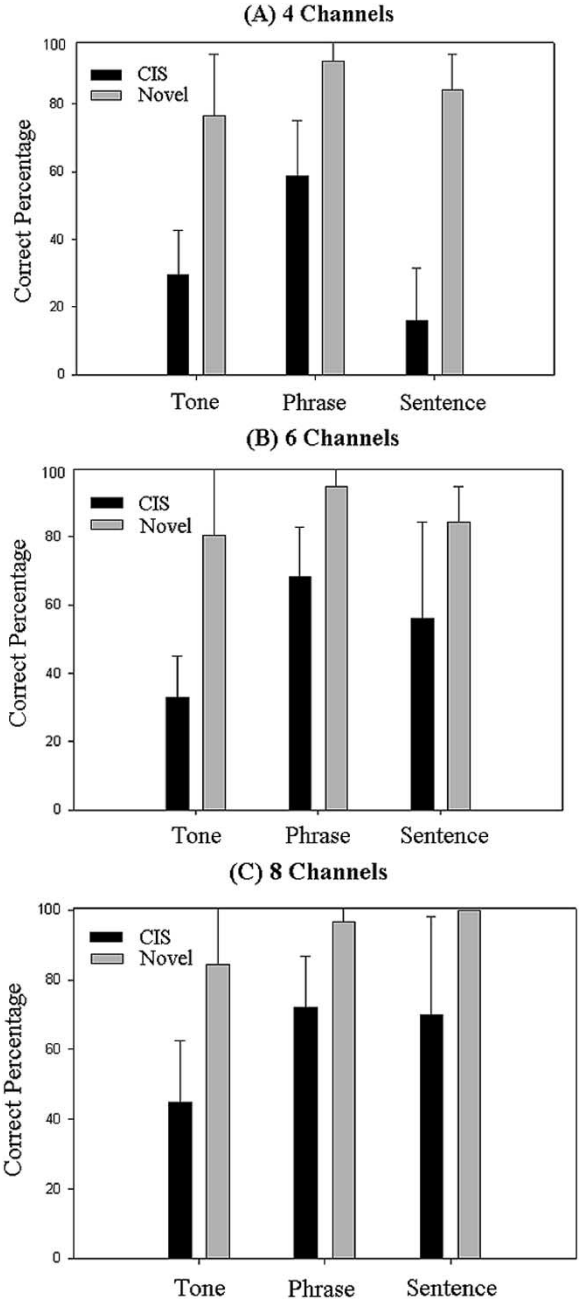

Fig. 8. In this figure, test results are re-plotted for comparison of the performance of the two strategies with 4, 6, and 8 channels, in order to show more clearly the improvement of the novel strategy. Black bars represent the test results of the CIS strategy (marked with "CIS"), and gray bars indicate the test results of the novel strategy (marked with "Novel"). The error bars symbolize the standard deviations.

information. The novel speech-processing strategy that we developed incorporates both envelope and tonal information by explicitly extracting F_0 from speech signals, and uses the F_0 pattern to dynamically modulate the center frequency of stimulus pulses. Thus, the novel processor transmits a greater amount of spectral information than the CIS processor, because the former utilizes an additional channel of frequency information contained in F_0 [see (2)]. This proves to be more effective in the perception of tones, phrases and sentences in Chinese speech.

Results of acoustic simulation showed that the CIS processor with 6 and 8 channels achieved at best a 40% correct rate for tone perception, and a 60%–70% correct rate for phrase and sentence perceptions in normal hearing subjects [Fig. 8(b) and (c)]. The level of performance of the simulated CIS strategy in this study was lower than those in the other acoustic simulation results [15], [16], [41]. This discrepancy may be attributed to a number of differences in the design of the present simulation study with previous ones. First, we used sinusoidal waves in the simulation, while the previous studies used noise carriers. Second, unlike the stimuli used in [16], we found no correlation between tone patterns and vowel duration in our stimuli. Thus, the potentially potent envelope cue in vowel duration could contribute to the difference in performance between the present and previous studies. And thirdly, the experimental protocols designed in this study tended to give subjects a shorter training (about 5 min) than in the previous studies. However, the low level of performance by the CIS strategy was consistent with the tone recognition results in cochlear implant users reported in [18], [33], [34], and [39].

Nevertheless, the present study clearly demonstrated, in the same subject pool, that the novel processing strategy produced significantly better speech perception performance than the CIS strategy. We noted, in particular, that the largest performance improvement was with the 4-channel novel processor in sentence recognition [Fig. 8(a)]. This result suggests that a device with fewer channels [38], if incorporated with the novel strategy, could theoretically perform adequately for tonal language perception. In the design and manufacture of a cochlear implant device, the number of channels in speech processing is a significant technical factor that relates to the cost, power consumption, the complexity of internal electronics, and the external communication interface. The present results are encouraging in that it is feasible to design an inexpensive, yet effective cochlear implant device for deaf patients who speak tonal languages.

The coding of F_0 information in the novel processor is in a way similar to that in the early Nucleus device. However, frequency modulation in the novel processor is limited to a range defined by F_0 and centered at a fixed frequency. In addition, the novel processor combines the technique of CIS that has been proven to be effective for English and German languages by cochlear implant users [29], [35]. This should facilitate the real-time implementation of the novel processor in currently available devices, such as the Clarion ®, with some modification to allow cycle-to-cycle variation of stimulation frequency. Although this and other studies [4], [28], [32] have shown that variable stimulation frequency can convey a certain amount of information on the fundamental frequency F_0 in the speech signal, it is still different from the way that the auditory nerve system encodes sound frequency information. In addition, the extent to which frequency modulation can be implemented is limited because of the refractoriness of auditory nerve response to repetitive stimulation [11]. These issues related to real-time implementation of the novel processor await future studies.

ACKNOWLEDGMENT

The authors would like to express their appreciation to Dr. Loeb for reading the manuscript and the anonymous reviewers for their comments. This manuscript was based on the doctoral dissertation of one of the authors, K.B. Nie, while at Tsinghua University, Beijing, China.

REFERENCES

[1] S. U. Ay, F. G. Zeng, and B. J. Sheu, "Hearing with bionic ears," *IEEE Circuits Devices Mag.*, vol. 13, pp. 18–23, Mar. 1997.

[2] P. J. Blamey, "An acoustic model of multiple-channel cochlear implant," *J. Acoust. Soc. Amer.*, vol. 76, no. 1, pp. 97–103, 1984.

[3] C. Boex, M. Pelizzone, and Montandon, "Speech recognition with a CIS strategy for the ineraid multichannel cochlear implant," *Amer. J. Otol.*, vol. 17, no. 1, pp. 61–68, 1996.

[4] P. A. Busby, Y. C. Tong, and G. M. Clark, "Electrode position, repetition rate, and speech perception by early and late-deafened cochlear implant patients," *J. Acoust. Soc. Amer.*, vol. 93, no. 2, pp. 1058–1067, 1993.

[5] V. Ciocca, A. L. Francis, R. Aisha, and L. Wong, "The perception of cantonese lexical tones by early-deafened cochlear implantees," *J. Acoust. Soc. Amer.*, vol. 111, pp. 2250–2256, 2002.

[6] M. F. Dorman *et al.*, "The recognition of sentences in noise by normal-hearing listeners using simulation of cochlear implant signal processors with 6–20 channels," *J. Acoust. Soc. Amer.*, vol. 104, no. 6, pp. 3583–3585, 1998.

[7] M. F. Dorman and P. C. Loizou, "Mechanisms of vowel recognition for ineraid patients fit with continuous interleaved sampling processor," *J. Acoust. Soc. Amer.*, vol. 102, no. 1, pp. 581–587, 1997.

[8] M. F. Dorman, P. C. Loizou, and D. Rainey, "Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs," *J. Acoust. Soc. Amer.*, vol. 102, no. 4, pp. 2403–2411, 1997.

[9] ——, "Simulating the effect of cochlear-implant insertion depth on speech understanding," *J. Acoust. Soc. Amer.*, vol. 102, no. 5, pp. 2993–2996, 1997.

[10] M. F. Dorman and P. C. Loizou, "The identification of consonants and vowels by cochlear implants patients using a 6-channel CIS processor and by normal hearing listeners using simulations of processors with two to nine channels," *Ear and Hearing*, vol. 19, pp. 162–166, 1998.

[11] D. K. Eddington, W. H. Dobelle, D. E. Brackmann, M. G. Mladejovsky, and J. L. Parkin, "Auditory prostheses research with multiple channel intracochlear stimulation in man," *Ann. Otol., Rhinol. and Laryngol.*, vol. 87, pp. 1–39, 1978.

[12] ——, "Place and periodicity pitch by stimulation of multiple scala tympani electrodes in deaf volunteers," *Trans. Amer. Soc. Artific. Internal Organs*, vol. 24, pp. 1–5, 1978.

[13] A. Faulkner, S. Rosen, and C. Smith, "Effects of the salience of pitch and periodicity information on the intelligibility of four-channel vocoded speech: Implication for cochlear implants," *J. Acoust. Soc. Amer.*, vol. 108, no. 4, pp. 1877–1887, 2000.

[14] J. H. M. Frijins, J. J. Briaire, and J. Grote, "The importance of human cochlear anatomy for the results of modiolus-hugging multichannel cochlear implants," *Otol. & Neurotol.*, vol. 22, pp. 340–349, 2001.

[15] Q. J. Fu, F. G. Zeng, and R. V. Shannon, "Importance of tonal envelope cues in Chinese speech recognition," *J. Acoust. Soc. Amer.*, vol. 104, no. 1, pp. 505–510, 1998.

[16] Q. J. Fu and F. G. Zeng, "Identification of temporal envelope cues in Chinese tone recognition," *Asia Pacific J. Speech, Lang., Hearing*, vol. 5, pp. 45–57, 2000.

[17] L. Geurts and J. Wouter, "Coding of the fundamental frequency in continuous interleaved sampling processor for cochlear implants," *J. Acoust. Soc. Amer.*, vol. 109, no. 2, pp. 713–726, 2001.

[18] T. S. Huang, N. M. Wang, and S. Y. Liu, "Tone perception of Mandarin-speaking postlingually deaf implantees using the nucleus 22-channel cochlear mini system," *Ann. Otol., Rhinol. and Laryngol.*, vol. 166, pp. 294–298, 1995.
[19] G. E. Loeb, "Cochlear prosthetics," *Ann. Rev. Neurosci.*, vol. 13, pp. 357–371, 1990.
[20] G. E. Loeb and B. S. Wilson, "Prosthetics, sensory systems," in *The Handbook of Brain Theory and Neural Networks*, 2nd ed, M. A. Arbib, Ed. Cambridge, MA: MIT Press, 2003, pp. 926–929.
[21] P. C. Loizou, "Mimicking the human ear," *IEEE Signal Processing Mag.*, vol. 15, pp. 101–130, Sept. 1998.
[22] ——, "An introduction to cochlear implant," *IEEE Eng. Med. Biol. Mag.*, vol. 18, pp. 32–42, Jan./Feb. 1999.
[23] P. C. Loizou, M. F. Dorman, and V. Powell, "The recognition of vowels produced by men, women, boys and girls by cochlear implant patients using a six-channel CIS processor," *J. Acoust. Soc. Amer.*, vol. 103, no. 2, pp. 1141–1149, 1998.
[24] H. J. McDermott, C. M. McKay, and A. E. Vandali, "A new portable sound processor for the university of melbourne/nucleus limited multi-electrode cochlear implant," *J. Acoust. Soc. Amer.*, vol. 91, no. 6, pp. 3367–3371, 1992.
[25] H. J. McDermott, A. E. Vandali, and R. J. Van Hoesel, "A portable programmable digital sound processor for cochlear implant research," *IEEE Trans. Rehab. Eng.*, vol. 1, pp. 94–100, June 1993.
[26] K. B. Nie, N. Lan, and S. K. Gao, "A speech processing strategy of cochlear implants based on tonal information of Chinese language," in *Proc. 20th Ann. Int. Conf. IEEE/EMBS*, Hong Kong, 1998, pp. 3154–3157.
[27] ——, "Acoustical simulation of speech processing strategy for cochlear implants," *J. Beijing Biomed. Eng.*, 1999.
[28] K. B. Nie, J. Liu, and S. K. Gao, "A speech processing strategy for cochlear implants based on tonal information of Chinese language," *Chinese J. Biomed. Eng.*, vol. 20, pp. 242–247, 2001.
[29] R. V. Shannon, F. G. Zeng, and V. Kamath, "Speech recognition with primarily temporal cues," *Science*, vol. 270, no. 13, pp. 303–304, 1995.
[30] M. W. Skinner, L. K. Holden, T. A. Holden, R. C. Dowell, P. M. Seligman, J. A. Brimacombe, and A. L. Beiter, "Performance of postlinguistically deaf adults with the wearable speech processor (WSP III) and mini speech processor (MSP) of the nucleus multi-electrode cochlear implant," *Ear and Hearing*, vol. 12, no. 1, pp. 3–22, 1991.
[31] M. W. Skineer *et al.*, "Identification of speech by cochlear implant recipients with the multipeak (MPEAK) and spetral peak (SPEAK) speech coding strategies 1. vowels," *Ear and Hearing*, vol. 17, no. 3, pp. 182–197, 1996.
[32] B. Townshend, N. Cotter, and D. V. Compernoll, "Pitch perception by cochlear implant subjects," *J. Acoust. Soc. Amer.*, vol. 82, no. 1, pp. 106–115, 1987.
[33] C. G. Wei, K. L. Cao, Z. Z. Wang, and F. G. Zeng, "Rate discrimination and tone recognition in mandarin-speaking cochlear-implant listeners," *Chin. J. Otorhinolaryngol.*, vol. 34, pp. 84–88, 1999.
[34] W. I. Wei, R. Wong, Y. Hui, D. K. Au, B. Y. Wong, W. K. Ho, A. Tsang, P. Kung, and E. Chung, "Chinese tonal language rehabilitation following cochlear implantation in children," *Acta Otolaryngol.*, vol. 120, pp. 218–221, 2000.
[35] B. S. Wilson, D. T. Lawson, and C. C. Finley, "New processing strategies in cochlear implantation," *Amer. J. Otol.*, vol. 16, pp. 668–675, 1995.
[36] B. S. Wilson, C. C. Finley, and D. T. Lawson, "Better speech recognition with cochlear implants," *Nature*, vol. 352, no. 18, pp. 236–238, July 1991.
[37] ——, "Design and evaluation of continuous interleaved sampling (CIS) processing strategy for multichannel cochlear implants," *J. Rehab. Res. Devel.*, vol. 30, pp. 110–116, 1993.
[38] B. S. Wilson, S. Rebscher, F. G. Zeng, R. V. Shannon, G. E. Loeb, D. T. Lawson, and M. Zerbi, "Design for an inexpensive but effective cochlear implant," *Otolaryngol. Head and Neck Surgery*, vol. 118, pp. 235–241, 1998.
[39] S. A. Xu, R. C. Dowell, and G. M. Clark, "Results from Chinese and English in a multichannel cochlear implant patient," *Ann. Otol., Rhinol. and Laryngol.*, vol. 96, pp. 126–127, 1987.
[40] L. Xu and B. E. Pfingst, "Relative importance of envelope and fine structure for tonal-speech perception as revealed by auditory chimera," in *Proc. 2001 Conf. Implantable Auditory Prostheses*, Asilomar, CA, 2001, p. 167.
[41] L. Xu, Y. J. Tsai, and B. E. Pfingst, "Features of stimulation affecting tonal-speech perception: Implication for cochlear prostheses," *J. Acoust. Soc. Amer.*, vol. 112, pp. 247–258, 2002.
[42] X. J. Yang and H. S. Chi, *Speech Signal Processing*. Beijjing: The Electronic Industry Press of China, 1995.
[43] T. R. Yao, *Digital Speech Processing*. Huazong, China: Huazhong Sci. Technol. Univ. Press, 1992.
[44] F. G. Zeng, "Cochlear implants in China," *Audiology*, vol. 34, pp. 61–75, 1995.
[45] F. G. Zeng, K. L. Cao, and Z. Z. Wang, "Progress in cochlear implants," *Chin. J. Otolaryngol.*, vol. 33, no. 2, pp. 123–125, 1998.

N. Lan, photograph and biography not available at the time of publication.

K. B. Nie, photograph and biography not available at the time of publication.

S. K. Gao, photograph and biography not available at the time of publication.

F. G. Zeng, photograph and biography not available at the time of publication.