ISCA Archive
http://www.isca-speech.org/archive

International Symposium on Chinese Spoken Language Processing (ISCSLP 2004)
Hong Kong
December 15-18, 2004

FREQUENCY CHANNELS TO CHINESE SPEECH RECOGNITION

Xin Luo *

Department of Electrical Engineering and Information Science
University of Science and Technology of China
Hefei, Anhui, 230027

Qian-Jie Fu †

Department of Auditory Implants and Perception
House Ear Institute
Los Angeles, California, 90057

ABSTRACT

Studies have revealed near perfect speech recognition with primarily temporal envelope cues and severely degraded spectral cues. Among different types of temporal envelope cues, periodicity fluctuation cues have been found to significantly improve Chinese tone and sentence recognition, while the contributions of periodicity fluctuation cues in individual frequency channels to Chinese speech recognition have not been clearly stated. In order to make periodicity fluctuation cues available in different frequency regions, the present study employed different low-pass cutoff frequencies for the temporal envelope detectors in different channels of a four-channel noise-band cochlear implant simulation. Chinese tone and vowel recognition scores were measured for six native Chinese normal hearing listeners under six low-pass cutoff frequency combinations: all 50 Hz in four channels (all-50 Hz), all 500 Hz in four channels (all-500 Hz), and 500 Hz in one of the four channels while 50 Hz in the other three channels (ch1-500 Hz, ch2-500 Hz, et al.). The results showed that the ch4-500 Hz condition produced the highest Chinese tone recognition among the four single-channel-500 Hz conditions, and was the only condition whose tone recognition was similar to that of the all-500 Hz condition and was significantly higher than that of the all-50 Hz condition. Chinese vowel recognition was not significantly affected by different cutoff frequency combinations. These results suggest that delivering periodicity fluctuation cues in higher frequency channels might be more important and efficient in enhancing Chinese tone recognition for cochlear implant patients.

1. INTRODUCTION

Cochlear implants generally divide acoustic signals into several frequency bands, extract the temporal envelope information from each band, convert the temporal envelope amplitudes into electric currents and deliver the electric currents to appropriate electrodes situated within the cochlea (e.g., [1, 2]). To recreate the tonotopic distribution of activity within the normal cochlea, the amplitude envelopes from low frequency bands are delivered to the electrodes located in the apical region and the amplitude envelopes from high frequency bands are delivered to the electrodes located in the basal region. While this electrical stimulation strategy provides primarily temporal envelope cues and gross spectral cues to profoundly deaf patients to restore their speech understanding, many other cues are not well transmitted, including the fundamental frequency and the temporal fine structure above 500 Hz within each spectral channel.

Although spectral cues in speech sounds have been thought required for speech recognition, many studies have shown near perfect speech recognition with primarily temporal envelope cues and severely degraded spectral cues. Shannon et al. [3] systematically manipulated available spectral and temporal cues using a noise-band simulation of the electrical stimulation of cochlear implants, and found that 500 Hz temporal envelope cues from only four broadband frequency bands were sufficient for the recognition of English speech by normal hearing listeners. In the noise-band simulation, temporal envelopes were extracted from broadband frequency bands and were used to modulate noises of the same bandwidths, which were then added together to produce the output speech. Fu et al. [4] similarly found about 90 % correct Chinese sentence recognition by normal hearing listeners listening to the same four-channel noise-band simulation. Besides these simulation results, applications of cochlear implants also have demonstrated moderate speech recognition performances with primarily temporal envelope patterns transmitted by the prosthesis devices (e.g., [1]).

Based on the frequency range of temporal waveforms, temporal envelope cues can be categorized into three types: amplitude envelope (< 50 Hz), periodicity fluctuation (50 - 500 Hz), and fine structure (> 500 Hz) [5]. Each of the three cues contains specific information for speech recognition. For example, amplitude envelope cues not only provide rough distributions of spectral energy but also provide

*The first author performed the work while at House Ear Institute

†The research was supported by grant R01-DC04993 from NIDCD

indications of syllables' tonal patterns, while periodicity fluctuation cues cover human's pitch range, and therefore transmit pitch information of speech sounds (e.g., [6]). Although Shannon et al. [3] observed no significant differences between English speech recognition results obtained with 50 and 500 Hz low-pass temporal envelopes, Fu et al. [4] found that increasing the cutoff frequencies of the temporal envelope filters from 50 to 500 Hz had a significant effect on Chinese tone and sentence recognition in a 4-channel cochlear implant simulation, which indicated the contributions of periodicity fluctuation cues to Chinese speech recognition.

In the previous studies, low-pass temporal envelope filters in individual frequency channels had the same cutoff frequencies (either 50 or 500 Hz), so the contributions of periodicity fluctuation cues in individual frequency channels to Chinese speech recognition were not clearly demonstrated. In the present study, different low-pass cutoff frequencies were used for the temporal envelope detectors in different channels of a 4-channel noise-band cochlear implant simulation, to make periodicity fluctuation cues available in different frequency regions. Six combinations of low-pass cutoff frequencies for the four channels were tested to investigate the effects of periodicity fluctuation cues in individual frequency channels on Chinese speech recognition. For the baseline condition (all-50 Hz) where periodicity fluctuation cues were not available in any of the four frequency regions, 50 Hz low-pass filters were used for all of the four channels to extract the 50 Hz amplitude envelopes only. On the other extreme, 500 Hz low-pass filters were used for all of the four channels in the "best" condition (all-500 Hz) so that periodicity fluctuation cues were available in all of the four frequency regions. In each one of the rest four conditions (ch1-500 Hz, ch2-500 Hz, et al.), periodicity fluctuation cues were only available in one of the four channels because 500 Hz low-pass filters were only used for the corresponding one channel, while 50 Hz low-pass filters were used for the other three channels. Both Chinese tone and vowel recognition scores were measured for 6 Chinese-speaking normal hearing subjects listening to 4-channel noise-band speech with periodicity fluctuation cues available in different frequency channels.

2. METHODS

2.1. Subjects

Six young adult native Chinese-speaking listeners (3 males, 3 females) participated in this study. All subjects were normal hearing and had pure-tone thresholds better than 20 dB HL at octave frequencies from 125 Hz to 8000 Hz in both ears.

2.2. Stimuli and Speech Processing

The Chinese vowel stimuli used in the present study were derived from the 'Chinese Standard Database' [7]. Five male and five female speakers each produced 4 tones for 6 Mandarin Chinese single-vowel syllables (/a/, /o/, /e/, /i/, /u/, /ü/), resulting in a total of 240 vowel tokens. These vowel stimuli were used for measuring both Chinese vowel and tone recognition. These stimuli were digitized using a 16-bit A/D converter at a 16-kHz sampling rate without high frequency pre-emphasis.

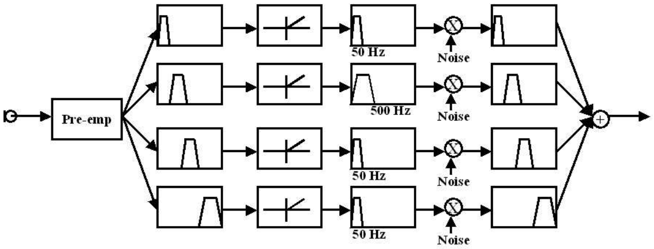

Fig. 1. Block diagram of the CIS simulation using different low-pass cutoff frequencies for temporal envelope detectors in individual frequency channels (Ch2-500 Hz condition: periodicity fluctuation cues only available in channel 2).

All speech stimuli were processed using a 4-channel noise-band acoustic simulation of a cochlear implant fitted with the Continuous Interleaved Sampling (CIS) strategy [1], as shown in figure 1. After pre-emphasis (1st-order Butterworth high-pass filter at 1200 Hz), the input speech signal was divided into 4 frequency bands (overall frequency range was between 300 and 6000 Hz). The corner frequencies of the analysis bands were determined according to Greenwood's formula [8]; all analysis filters were 4th-order Butterworth band-pass filters. The temporal envelope from each analysis band was extracted by half-wave rectification and low-pass filtering (4th-order Butterworth low-pass filters), the low-pass cutoff frequencies of the temporal envelope detectors in individual frequency channels for different conditions were listed in table 1. The extracted temporal envelopes were used to modulate wide-band noise, which were then band-pass filtered by filters with the same passbands as the analysis filters. The output speech was the sum of these modulated noise bands.

2.3. Procedure

Tone recognition tests were conducted using a 4-alternative, forced-choice (4-AFC) task, while vowel recognition tests were conducted using a 6-AFC task; no feedback was provided. Subjects were seated in a double-walled sound-treated

Combination	LP cutoff freqs of temporal envelope filters in channels			
Conditions	Ch 1 (Hz)	Ch 2 (Hz)	Ch 3 (Hz)	Ch 4 (Hz)
All-50	50	50	50	50
Ch1-500	500	50	50	50
Ch2-500	50	500	50	50
Ch3-500	50	50	500	50
Ch4-500	50	50	50	500
All-500	500	500	500	500

Table 1. Low-pass cutoff frequencies of temporal envelope detectors in individual frequency channels for different conditions

booth and listened to the stimuli presented in free-field over a single loudspeaker (Tannoy Reveal) at 65 dBA. For each recognition task, the test order of speech processing conditions was randomized and counterbalanced across subjects.

3. RESULTS

Figure 2 shows Chinese tone and vowel recognition scores obtained with the six temporal envelope cutoff frequency combinations. Chinese tone recognition was significantly affected by the availability of periodicity fluctuation cues in different frequency channels [one-way ANOVA: $F(5, 30) = 3.39, P = 0.02$], while Chinese vowel recognition was not [$F(5, 30) = 0.13, P = 0.99$]. The inclusion of periodicity fluctuation cues in all of the four channels improved the average tone recognition score from 60.90 to 67.43 % correct, but did not change the average vowel recognition score (64 % correct). These results were consistent with previous studies [4].

Among the four conditions where periodicity fluctuation cues were only available in single channel, the ch4-500 Hz condition produced similar tone recognition scores as the all-500 Hz condition, and was the only condition that had significantly higher tone recognition scores than the all-50 Hz condition [Student's t-test: $t(10) = 3.95, p = 0.01$]. For vowel recognition, none of the four conditions yielded significantly different results compared to the all-50 Hz condition.

4. DISCUSSION

The fact that the ch4-500 Hz condition produced significantly higher Chinese tone recognition scores than the other 3 single-channel-500 Hz conditions suggest that periodicity fluctuation cues in the 4th frequency channel contributed the most to Chinese tone recognition. The reason of this result may be analyzed in two aspects: waveform characteristics of periodicity fluctuations in different frequency channels and sensitivities of different cochlear positions of normal hearing listeners to periodicity fluctuations.

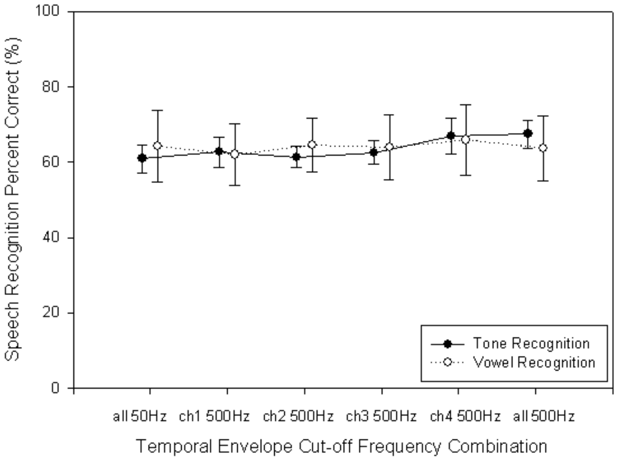

Fig. 2. Chinese speech recognition as a function of the different combinations of low-pass temporal envelope cutoff frequencies in individual channels

Even if the same temporal envelope detector (half-wave rectifier and 4th-order Butterworth low-pass filter at 500 Hz) was used to extract periodicity fluctuations in different frequency channels, there still exist some apparent differences between the periodicity fluctuation waveforms in different frequency channels. Seen from the example in figure 3, the relative modulation depths of the 500 Hz temporal envelopes are similar for the four channels, but channel 4 has the simplest periodicity fluctuation waveforms, which might benefit the detection of pitch and its changes [9].

Higher Chinese tone recognition scores may also come from higher sensitivities to periodicity fluctuation cues. For cochlear implant patients, studies have not found big differences or consistent patterns in amplitude modulation detection thresholds across different electrodes (i.e., different frequency regions of cochlear) [10, 11]. But for normal hearing listeners, different bandwidths of the critical bands in different frequency regions may cause different processing mechanisms for the periodicity fluctuations. In channel 4, within-critical-band processing is enough for 500 Hz periodicity fluctuation cues; while in channel 1, across-critical-band processing should be employed. These different mechanisms might make periodicity fluctuations easier to be detected in higher frequency channels. This hypothesis could be further tested by comparing the amplitude modulation detection thresholds of the same modulation waveform carried by narrow-band noises in different frequency regions.

The results of the present study suggest that periodicity fluctuation cues in the 4th frequency channel might be the most important to Chinese tone recognition by normal hear-

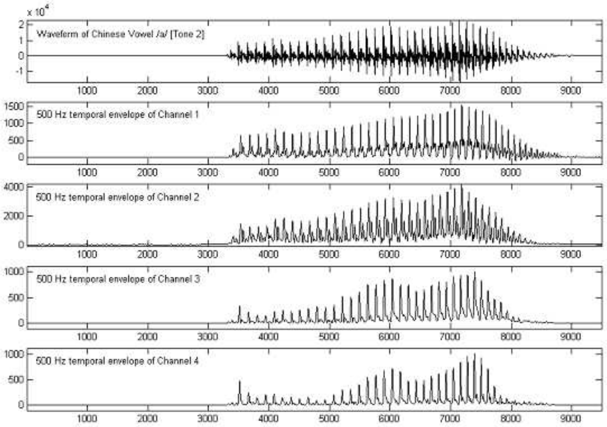

Fig. 3. Example of the 500 Hz temporal envelopes in 4 frequency channels. The first panel shows the original waveform of a Chinese vowel /a/ (tone 2). The following panels show the 500 Hz temporal envelopes of the Chinese vowel /a/ (tone 2) in channel 1, 2, 3, and 4, respectively.

ing listeners. A possible indication for Chinese cochlear implant design is that when overall stimulation pulse rate is limited, it is better to arrange higher pulse rates for basal electrodes so that periodicity fluctuation cues can be better transmitted in high frequency channels.

5. CONCLUSIONS

Chinese tone recognition was significantly affected by the availability of periodicity fluctuation cues in different frequency channels, while Chinese vowel recognition was not significantly affected. In a 4-channel noise-band cochlear implant simulation, among the four conditions where 500 Hz temporal envelopes were available in only one of the four channels and 50 Hz temporal envelopes were available in the other three channels, the ch4-500 Hz condition produced the highest Chinese tone recognition and was the only condition whose tone recognition was similar to that of the all-500 Hz condition and was significantly higher than that of the all-50 Hz condition. These results suggest that delivering periodicity fluctuation cues in higher frequency channels might be more important and efficient in enhancing Chinese tone recognition for cochlear implant patients.

6. REFERENCES

[1] B. S. Wilson, C. C. Finley, D. T. Lawson, et al., “Better speech recognition with cochlear implants,” *Nature (London)*, vol. 352, pp. 236–238, 1991.

[2] M. W. Skinner, G. M. Clark, and L. A. Whitford, “Evaluation of a new Spectral Peak coding strategy for the Nucleus 22 channel cochlear implant system,” *American Journal of Otolaryngology*, vol. 15, suppl. 2, pp. 15–27, 1994.

[3] R. V. Shannon, F.-G. Zeng, V. Kamath, J. Wygonski, and M. Ekelid, “Speech recognition with primarily temporal cues,” *Science*, vol. 270, pp. 303–304, 1995.

[4] Q.-J. Fu, F.-G. Zeng, R. V. Shannon, and S. D. Soli, “Importance of tonal envelope cues in Chinese speech recognition,” *Journal of the Acoustical Society of America*, vol. 104, pp. 505–510, 1998.

[5] S. Rosen, “Temporal informaiton in speech: acoustic, auditory and linguistics aspects,” *Philos. Trans. R. Soc. London*, vol. Ser. B 336, pp. 367–373, 1992.

[6] Q.-J. Fu and F.-G. Zeng, “Identification of temporal envelope cues in Chinese tone recognition,” *Asia Pacific Journal of Speech, Language, and Hearing*, vol. 5, pp. 45–57, 2000.

[7] R.-H. Wang, “The standard Chinese database,” 1993, University of Science and Technology of China, internal materials.

[8] D. D. Greenwood, “A cochlear frequency-position function for several species – 29 years later,” *Journal of the Acoustical Society of America*, vol. 87, pp. 2592–2605, 1990.

[9] T. Green, S. Rosen, and A. Faulkner, “Enhancing temporal cues to voice pitch available through cochlear implant speech processors,” *Assoc. Res. Otolaryngol. Abstr.*, vol. 26, pp. 198, 2003.

[10] L. M. Richardson, P. A. Busby, and G. M. Clark, “Modulation detection interference in cochlear implant subjects,” *Journal of the Acoustical Society of America*, vol. 104, pp. 442–452, 1998.

[11] R. V. Shannon, “Temporal modulation transfer functions in patients with cochlear implants,” *Journal of the Acoustical Society of America*, vol. 91, pp. 2156–2164, 1992.