Effective use of the spectral information in speech processing of cochlear implant*

GUAN Tian[1], GONG Qin[2] and YE Datian[2]**

(1. Joint Research Centre for Biomedical Engineering, Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong, China; 2. Research Centre of Biomedical Engineering, Graduate School at Shenzhen, Tsinghua University, Shenzhen 518055, China)

Accepted on September 14, 2006

Abstract Based on the results of the acoustic research on Mandarin, two novel algorithms using the spectral information in speech processing of cochlear implants are purposed and certified effectively by the spectral information of tonal language in acoustic simulation experiments. We bring forward novel algorithms conveying the spectral information based on the choices of the frequency bands. These new algorithms can not only improve the speech recognition ability of cochlear implant users in the noisy environments but also reduce the complexity of computing and the memory occupied, and make it more suitable to be carried out in clinical practice.

Keywords: **cochlear implant, speech signal processing, signal processing.**

Cochlear implant (CI) is the only available medical device to restore hearing ability to totally deaf people by extracting the temporal speech signal envelope which is typically encoded to amplitude modulate a fixed-rate electric-stimulus pulse. More than 1600 persons have achieved modern multi-electrode CIs which successfully rebuild their auditory sense in China. Although the modern multi-channel devices produce speech recognition scores around 75% for sentences in quiet, the ability of most CI users to understand speech in noisy environments remains quite poor[1—3], especially the ability of understanding tonal information which is important in Mandarin speech recognition because the tonality of a monosyllable is lexically meaningful[4—6].

Many investigators focused on building novel algorithms of speech processing which not only conveyed the temporal envelop cues but also conveyed spectral information to enhance the speech recognition abilities of CI users. Chen et al.[7] and Nie et al.[8] derived amplitude modulation (AM) and frequency modulation (FM) signals from sounds and conducted acoustic simulation experiments in normal-hearing subjects. They found that additional encoding of FM could significantly enhance English speech recognition in noise. Lan et al.[9] developed a novel algorithm by extracting and encoding both the envelopes of narrow-band signals and fundamental frequency (F0) of the speech signal. F0 was used to modulate the center frequency of sinusoidal waves in acoustic simulation experiments. This algorithm produced significant improvement in perception of Mandarin. All these researches proved that spectral information could significantly enhance the ability of speech perception in CI users. However, based on three aspects of the investigations of phonetics research we assumed that it had redundancy to transmit this kind of information in every channel. A more compact algorithm could be brought forward after reducing the redundancy of conveying spectral information.

First, the pipelines of conveying the tonal information of Mandarin had redundancy. Both temporal information and spectral information of speech signals contributed to the recognition of four Mandarin tones[10]. Many investigations isolated the spectral envelope cues and found that the temporal envelope information such as vowel duration and amplitude contours contributed to Mandarin tone recognition[11—14]. This contribution, while significant, was relatively weak when the spectral pitch information evoked by fundamental frequency and its harmonics were present[15]. So there were multiple pipelines of conveying the tonal information. And perfect tone recognition scores could be gained even if some of pipelines were isolated, so the pipelines of conveying the tonal information of Mandarin had redundancy.

* Supported by National Natural Science Foundation of China (Grant No. 30570472) and Hong Kong Innovation and Technology Fund (Grant No. GHS/053/04)

** To whom correspondence should be addressed. E-mail: yedt@sz.tsinghua.edu.cn

?1994-2018 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

Secondly, perfect tone recognition could be achieved by only extracting and encoding the temporal and spectral information ranging in low frequencies. Previous work found that perfect tone recognition could be achieved by either directly preserving the fundamental frequency with low-pass filtering at 300 Hz or indirectly by residual pitch derived from the harmonic structure which also ranged in low frequencies[16]. Therefore, maybe conveying the temporal and spectral information in low frequency bands was enough to gain perfect speech recognition.

Finally, the spectral information encoded by traditional algorithms in high frequency bands could hardly be apperceived in acoustic simulation experiments. Many investigators focused on how to extract and convey the spectral information to enhance the speech recognition ability[7—9]. Both of these two kinds of spectral information, such as F0 and FM were encoded in every frequency band of speech signals. But in high frequency bands, the transmission of F0 and FM made CI users hardly to apperceive the spectral information encoded by traditional algorithms. Because either in low frequency band or in high frequency band, the spectral information, which varied in low frequencies and ranged about 100 Hz, was used to modulate the central frequency of sinusoidal waves in acoustic simulation experiments. Therefore, in high frequency bands, the spectral information varying range was relative negligible with respect to the central frequency of sinusoidal waves corresponding to these bands (For example, the ratio of spectral information to central frequency of sinusoidal wave of 8-channel cochlear implant from the lowest to the highest frequency band were as follows: 47.4%, 28.4%, 17.5%, 11.1%, 7%, 4.5%, 3%, 1.9%.). And this made CI users hardly to apperceive the spectral information encoded by traditional algorithms in high frequency bands.

Based on these three aspects of investigations, we assumed that perfect speech recognition could be achieved when we extracted and encoded the temporal envelope and spectral information in the lower frequency bands but only extracted and encoded the temporal envelope in the higher frequency bands. This made the principle of frequency band choices in which the spectral information was extracted and encoded into the frequency modulation in the low frequency part (near the apex of cochlea) while the spectral information was not calculated or used in the high frequency part (near the base). The number of frequency ranges with the spectral information added from the apex (defined as the parameter S) was determined by the results of the acoustic simulation experiments.

We extracted and encoded the spectral information in two ways, i.e. Selective Fundamental Frequency Control (SFFC) algorithm and Selective Frequency Amplitude Modulation Encoding (SFAME) algorithm. Different speech materials in different environments were used to prove the validity of the algorithms in acoustic simulation experiments. The acoustic simulation experiments performed in this study expanded the research on the effect of FM information in Mandarin recognition both in white noise and mixed speech environment, which was not reported in literature[7,8] before. And these experiments also expanded the research on the effect of F0 in Mandarin speech recognition in multi-SNR white noise environment and in multi-TMR mixed speech environment, which has not been reported[9]. Similar results have been obtained from the use of different spectral information compared with the traditional algorithm—Continuous Interleaved Sampling (CIS)[17]. It also showed a great potential of using the spectral information to improve the speech recognition of cochlear implant users.

1 Algorithms

The CIS algorithm exists in all major clinical cochlear implant products. The input speech signal is first pre-emphasized above 1.2 kHz at 6 dB/Oct and then is separated into several bands (4, 6, 8, 12 etc.) by a bank of band-pass filters. The band of low frequency corresponds to the electrode which stimulates the top of the cochlea while the high frequency band corresponds to the electrode which stimulates the bottom of the cochlea. In each band, envelope signal can be obtained after a rectifier and a low-pass filter. In the electric-stimulus mode, the amplitudes of electric-stimulus pulse trains are modulated by the envelope signal extracted from the output of each frequency band. In the acoustic simulation model mode, the envelope is used to modulate the sine signals of the central frequency of the filter band and then resynthesize the modulation signals into simulation signals. Consequently, the envelope cues of speech signals from different frequency bands can be transmitted to the CI users.

Two algorithms have been adopted to extract and

?1994-2018 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

encode the spectral information, i.e. SFFC and SFAME, which avoid the disadvantage of current algorithms.

The SFFC algorithm is to extract and encode the fundamental frequency of the speech. This algorithm has two signal pathways, including the traditional envelope extraction like CIS algorithm and additional fundamental frequency processing. In one signal pathway which is similar to the standard CIS algorithm, after the phonetic signals have been pre-processed, the process of division of frequency bands and the extraction of the envelope are carried out; in the other signal pathway, the fundamental frequency is extracted by using lifting scheme and used to modulate the rate of electric-stimulus pulses under control of the principle of frequency band choice. In acoustic simulation model, when the simulation signals are re-synthesized, the fundamental frequency information is used under the control of the principle of frequency band choice as the modulation to the central frequency of the sine signals, which re-synthesizes the phonetics[9,18] (Fig. 1).

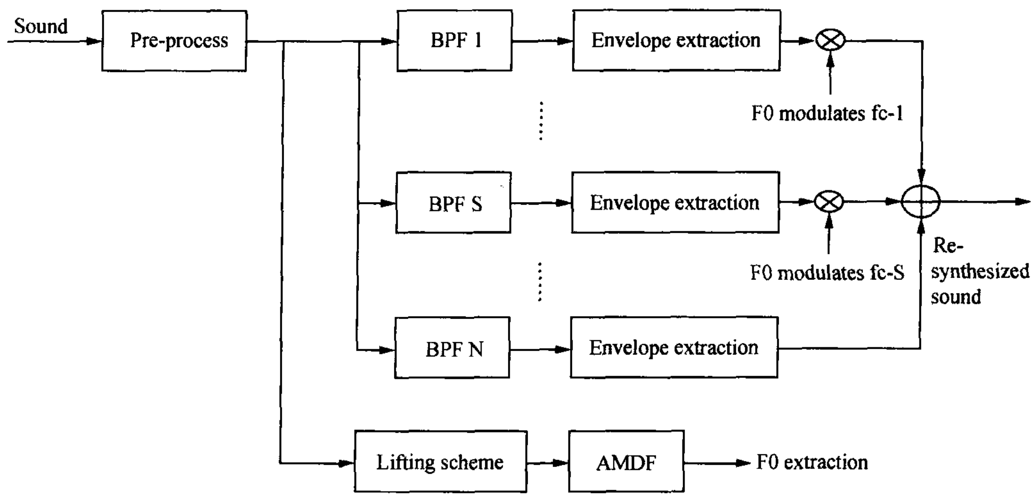

Fig. 1. Function block diagram of SFFC algorithm.

Similarly, SFAME algorithm uses the frequency modulation information to improve the speech recognition. This algorithm also has two signal pathways in each frequency band. In the first pathway, the traditional envelope extraction which is similar to the standard CIS algorithm is adopted like SFFC. But different from SFFC, SFAME adopts not the fundamental frequency as the spectral information conveyed, but the slowly varying frequency change information to modulate the pulse rate of the electrical simulation in the second signal pathway. By removing the central frequency of the sub-band signals and additionally limiting the frequency modulation's range and rate, SFAME algorithm transforms the fast-varying temporal fine structure into a slowly varying frequency modulation (FM) information. In acoustic simulation model, when the simulation signals are re-synthesized, the FM information is used under the control of the principle of frequency band choice as the modulation to the central frequency of the sine signals, which re-synthesizes the phonetics[7,8] (Fig. 2).

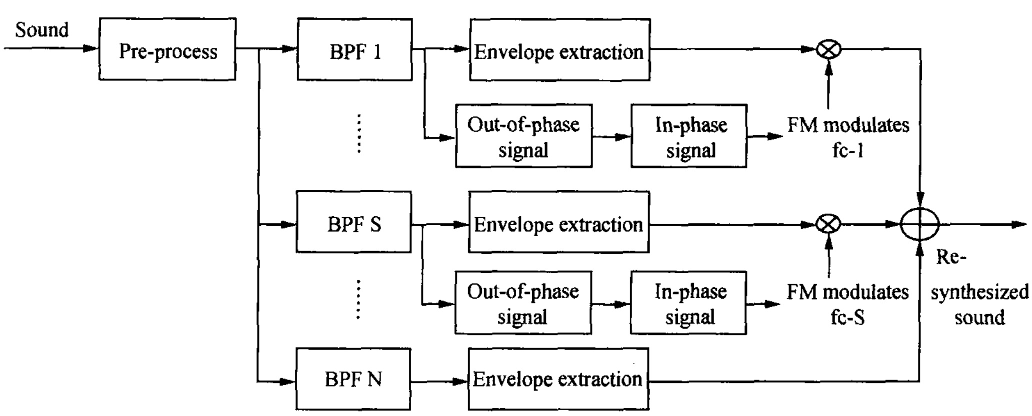

Fig. 2. Function block diagram of SFAME algorithm.

?1994-2018 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

2 Acoustic simulation experiment

Acoustic models of cochlear implant speech processors have been employed by many investigators to conduct listening experiments on normal-hearing subjects. We verified two speech processing algorithms in acoustic simulation experiments, trying to prove that the spectral information on the speech recognition of Mandarin, especially in the situation of white noise and mixed speech is more effective. The acoustic simulation experiments performed in this paper expanded the research on the effect of FM information in Mandarin recognition both in white noise and mixed speech environment, which was not referred to in literature[7,8]. And these experiments also expanded the research on the effect of F0 in Mandarin speech recognition in multi-SNR white noise environment and in multi-TMR mixed speech environment, which was not referred to in literature[9].

Twenty-four young native Mandarin speakers participated in this experiment. All subjects were reported with normal hearing. A quiet laboratory was selected to carry out the experiment. All the simulation sounds were presented via a Sennheiser HD457 headphone. The speech materials adopt vowels (close-set), words (open-set) and sentences (open-set). Two conditions were the white noise background and the mixed speech background (male-female overlapping) respectively. The rate of speech sampling was 16 kHz. The channel number was fixed at 8.

There were 100 questions in each experiment of vowels, words, sentences and mixed speeches. The SNR of the overlapping white noise (TMR in mixed speech experiments) was − 5 dB, 0 dB, 0 dB and 5 dB. The recognition rate of the vowels experiment was equal to the number of correct questions dividing by the total number of questions. While those of the words, sentences and mixed speech experiments were calculated using the number of correct keywords dividing by the total number of keywords.

3 Results

The results of the recognition rate based on various algorithms at different S levels using different language materials are shown in Figs. 3—5.

Tables 1 and 2 respectively present the data calculated by SFFC algorithm and SFAME algorithm in different situations based on AONVA analysis.

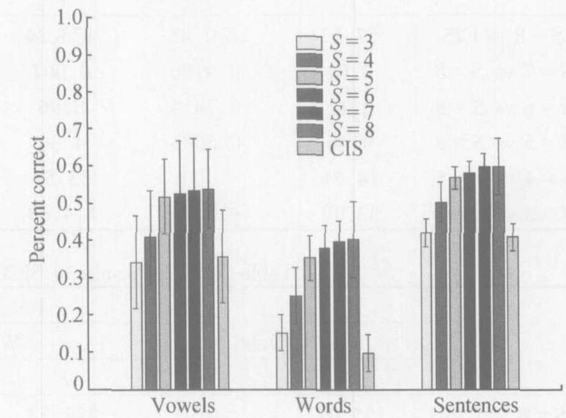

Fig. 3. Recognition rate of different language materials of SFFC compared with CIS in white noise background.

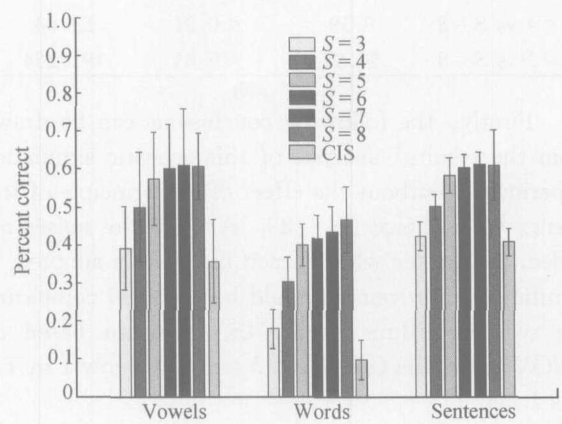

Fig. 4. Recognition rate of different language materials of SFAME compared with CIS in white noise background.

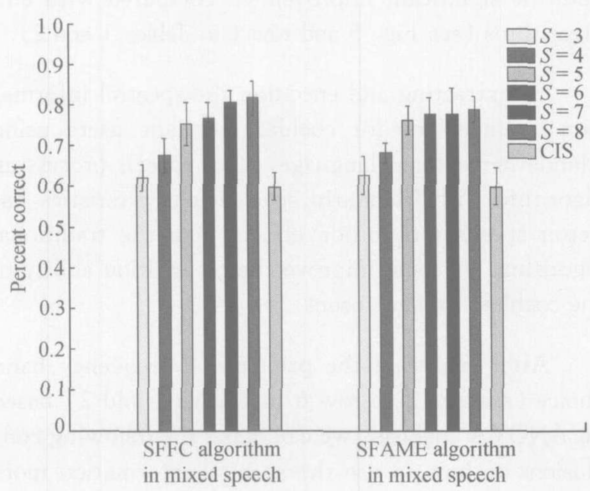

Fig. 5. Recognition rate of mixed speech of SFFC, SFAME compared with CIS.

?1994-2018 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

Table 1. The results of SFFC algorithm based on ANOVA analysis

	Results of SFFC algorithm analysis							
	Vowels		Words		Sentences		Mixed speech	
	F	*P*	*F*	*P*	*F*	*P*	*F*	*P*
$S=8$ vs CIS	57.63	<0.01	173.24	<0.01	114.08	<0.01	340.7	<0.01
$S=7$ vs $S=8$	0.01	0.9186	0.007	0.7959	0	0.971	1.38	0.2457
$S=6$ vs $S=8$	0.11	0.7419	0.96	0.3321	1.15	0.2892	17.85	<0.01
$S=5$ vs $S=8$	0.46	0.5005	4.36	0.0424	3.4	0.0717	21.5	<0.01
$S=4$ vs $S=8$	14.71	<0.01	33.59	<0.01	25.56	<0.01	131.34	<0.01
$S=3$ vs $S=8$	33.89	<0.01	114.66	<0.01	107.48	<0.01	319.56	<0.01

Table 2. The results of SFAME algorithm based on ANOVA analysis

	Results of SFAME algorithm analysis							
	Vowels		Words		Sentences		Mixed speech	
	F	*P*	*F*	*P*	*F*	*P*	*F*	*P*
$S=8$ vs CIS	155.97	<0.01	332.52	<0.01	92.79	<0.01	592.33	<0.01
$S=7$ vs $S=8$	0	0.9585	2.04	0.1599	0	0.9569	0.52	0.4756
$S=6$ vs $S=8$	0.05	0.8291	5.17	0.0277	1.14	0.7124	1.34	0.2532
$S=5$ vs $S=8$	1.06	0.3092	10.21	<0.01	1.63	0.2084	6.36	0.0152
$S=4$ vs $S=8$	9.59	<0.01	53.84	<0.01	29.74	<0.01	160.57	<0.01
$S=3$ vs $S=8$	56.52	<0.01	191.254	<0.01	83.75	<0.01	191.5	<0.01

Firstly, the following conclusions can be drawn from the results' analysis of this acoustic simulation experiments without the effect of the principle of frequency band choice ($S=8$). After white noises are added, no matter what speech material is adopted, a significant improvement could be obtained comparing the two algorithms with CIS algorithm based on ANOVA analysis (see Figs. 3 and 4 and row 1 in Tables 1 and 2).

When the speech materials are male-female overlapping (TMR = 5 dB), SFFC and SFAME also could gain the significant improvement compared with CIS algorithms (see Fig. 5 and row 1 in Tables 1 and 2).

So extracting and encoding the spectral information is important for cochlear implant users using Mandarin—a tonal language. The speech processing algorithm with Mandarin spectral characteristics has better speech recognition effects than the traditional algorithm. It could improve the recognition ability of the cochlear implant users.

After analyzing the principle of frequency band choice (see row 2 to row 6 in Tables 1 and 2) based on ANOVA analysis, we can make the following conclusions on how to use the spectral information more effectively.

(1) No matter the background noise is the white noise or the masking speech, and no matter the speech materials are vowels, words or sentences, the recognition rate of Mandarin continues to decrease with the reduction of S.

(2) When the channel number with spectral information is equal to or more than 5 ($S>=5$), there is no significant difference in the recognition rate between the algorithms with different S compared with the recognition rate without the effect of the principle of frequency band choice ($S=8$) in most situations.

(3) When the channel number S is less than 5 ($S<5$), there is significant decrease with the decrease of S.

Therefore, we can bring forward novel algorithms conveying the spectral information based on the choices of the frequency ranges. On the one hand, the extraction of the spectral information is similar to ways adopted by the original algorithm; on the other hand, the spectral information extracted is only used in a certain channel ($S=5$) in low frequency bands corresponding to the electrodes near the apex of the cochlea where the frequency coding mechanism—"temporal code" provides the applicable feasibility. These new algorithms of the spectral information can not only effectively improve the phonetic recognition ability of cochlear implant users in the noisy environment, but also reduce the complexity of

?1994-2018 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

computing and the memory occupied (for example, reduce by 37.5% of the amount of FM computing and reduce by 37.5% of the amount of F0 and FM transmission), so as to make it more suitable to be carried out in clinical practice.

References

1 Dorman M.F., Loizou P.C., Fitzke J. et al. The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear-implant signal processors with 6—20 channels. J. Acoust. Soc. Am., 1998, 104: 3583—3585.

2 Fu Q.J. and Shannon R.V. Phoneme recognition by cochlear implant users as a function of signal-to-noise ratio and nonlinear amplitude mapping. J. Acoust. Soc. Am., 1999, 106: L18—2.

3 Stickney G., Zeng F.G., Litovsky R. et al. Cochlear implant speech recognition with speech masker. J. Acoust. Soc. Am., 2004, 116: 1081—1091.

4 Wei C.G., Cao K.L., Wang Z.Z. et al. Rate discrimination and tone recognition in mandarin-speaking cochlear-implant listeners. Chin. J. Otorhinolaryngol., 1999, 34: 84—88.

5 Wei W.I., Wong R., Hui Y. et al. Chinese tonal language rehabilitation following cochlear implantation in children. Acta Otolaryngol., 2000, 120: 218—221.

6 Wu J.L. and Yang H.M. Speech perception of Mandarin Chinese speaking young children after cochlear implant use: effect of age at implantation. Int. J. Pediatr. Otorhinolaryngol., 2003, 67: 247—253.

7 Chen H.B. and Zeng F.G. Frequency modulation detection in cochlear implant subjects. J. Acoust. Soc. Am., 2004, 116: 2269—2277.

8 Nie K., Stickney G. and Zeng F.G. Encoding frequency modulation to improve cochlear implant performance in noise. IEEE Trans. Biomed. Eng., 2005, 52: 64—73.

9 Lan N., Nie K.B., Gao S.K. et al. A novel speech processing strategy incorporating tonal information for cochlear implants. IEEE Trans. Biomed. Eng., 2004, 51: 752—760.

10 Wei Ch.G., Cao K.L. and Zeng F.G. Mandarin tone recognition in cochlear-implant subjects. Hearing Research, 2004, 197: 87—95.

11 Blicher D. L., Diehl R. L. and Cohen L.B. Effects of syllable duration on the recognition of the Mandarin tone 2/tone 3 distinction: evidence for auditory enhancement. J. Phonetics, 1990, 18: 37—49.

12 Garding E., Kratochvil P., Svantesson J. O. et al. Tone 4 and tone 3 discrimination in modern standard Chinese. Lang. Speech, 1986, 29: 281—293.

13 Howie J. M. On the domain of tone in Mandarin. Phonetica., 1974, 30: 129—148.

14 Whalen D. H. and Xu Y. Information for Mandarin tones in the amplitude contour and in brief segments. Phonetica., 1992, 49: 25—47.

15 Lin M. C. Acoustic and perceptual cues for Mandarin tones. Chin. Linguist., 1988, 3: 182—193.

16 Liang Z. A. Auditory perceptual cues in Mandarin tones. Acta Physiol. Sin., 1963, 26: 85—91.

17 Wilson B.S., Lawson D.T., Zerbi M. et al. Design and evaluation of a Continuous Interleaved Sampling (CIS) processing strategy for multichannel cochlear implants. J. Rehabilitation Research and Development, 1993, 30(1): 110.

18 Luo X. and Fu Q.J. Importance of pitch and periodicity to Chinese-speaking cochlear implants patients. ICASSP, 2004, Ⅳ-1-4.

?1994-2018 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net