A Novel Speech Processing Algorithm for Cochlear Implant Based on Selective Fundamental Frequency Control

Tian Guan¹, Qin Gong¹, and Datian Ye²

¹ Department of Biomedical Engineering, Tsinghua Univ., Beijing 100084, P.R. China
² Research Center of Biomedical Engineering, Graduate School at Shenzhen, Tsinghua Univ., Shenzhen 518055, P.R. China

Abstract. The speech recognition ability of most CI users in noisy environment remains quite poor, especially for those who speak tonal language, such as Mandarin. Based on the results of the acoustic research on Mandarin, a novel algorithm using fundamental frequency was proposed, adding the principle of frequency bands selection. Using this principle F0 was encoded only in the selected frequency bands, which was confirmed effectively in acoustic simulation experiments with white noise or mixed speech environment. Compared with the traditional algorithm which only transmitted envelope cues, this novel strategy achieved remarkable improvement no matter adopting Mandarin vowels, words, sentences or mixed sentences. What’s more, the amount of transmission in this algorithm decreased to 62.5% compared with similar algorithms which transmit fundamental frequency in full channels.

1 Introduction

Cochlear implant is the available medical device to restore hearing ability to totally deaf people. Although the modern multi-channel devices produce speech recognition scores around 75% for sentences in quiet, the ability of most implant users to understand speech in noisy environment remains quite poor [1][2], especially for those who speak tonal language, such as Mandarin [3][4] in which pitch variations convey lexically different meanings.

The modern multi-channel devices mostly extract the temporal envelope, but discard the spectral information such as the pitch cues. Although the evidences show that the availability of pitch cues have little effect on performance in vowel and consonant recognition of English, [5] it is no doubt that for tonal language, such as Mandarin, pitch information makes significant contributions to speech perception, because in these languages pitch variations convey different lexical meanings. The same consonant-vowel syllable may mean totally different signification depending on fundamental frequency variations including flat, rising, falling-rising, and falling tone.

Many investigators focused on building novel algorithms of speech processing which not only conveyed the temporal envelop cues, but also conveyed the tonal
information. Lan N. (et al) [6] developed a novel algorithm by extracting and encoding both the envelope and fundamental frequency (F0) of speech signal. F0 was used to modulate the center frequency of sinusoidal waves in every channel in acoustic simulation experiments. This algorithm produced significant improvement in speech recognition ability of Mandarin. 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 envelope information and pitch information of speech signals contributed to the recognition of four Mandarin tones.[4] Many investigations found that the temporal envelope information such as vowel duration and amplitude contours contributed to Mandarin tone recognition. [7][8] This contribution, while significant, was relatively weak when the spectral pitch information evoked by fundamental frequency and its harmonics were present. [9] 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.

Second, perfect tone recognition could be achieved by only extracting and encoding the temporal and spectral information ranging in low frequency. 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 frequency [10]. Therefore, maybe conveying the temporal and tonal information in low frequency bands was enough to gain perfect speech recognition.

Last, the tonal information encoded by traditional algorithms in high frequency bands could hardly be apperceived. In lecture [6], the tonal information-F0 was used to modulate the central frequency of sinusoidal waves in acoustic simulation model. Therefore, in high frequency bands, the tonal information varying range was relative negligible with respect to the central frequency of sinusoidal waves corresponding to these bands (For example, the ratio of F0 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%. ).

Based on these theories, it is assumed that perfect speech recognition can be achieved when we encode the temporal envelope and tonal information in lower frequency bands but only encode the temporal envelope in the higher frequency bands. So we propose to encode the fundamental frequency more effectively in certain frequency bands selected by the principle frequency bands selection. This novel algorithm is called the selective fundamental frequency control (SFFC) algorithm. In the following section, we will first present the structure of SFFC in more details. Then the algorithm’s accuracy and efficiency will be verified using its acoustic simulation model in Section 3 and evaluate the results in Section 4.

2 Algorithm

The SFFC algorithm extracts and encodes the fundamental frequency of the speech signal. This algorithm has two signal pathways, including the traditional