THE CRITICAL ELEMENTS FOR CANTONESE SPEECH RECOGNITION – IMPLICATIONS FOR THE DESIGN OF DIGITAL HEARING AID AMPLIFICATION ALGORITHMS

Kevin C. P. Yuen[1]
Michael Tong[1]
Tan Lee[2]
Yuan Meng[2]
Sigfrid Soli[3]
Andrew van Hasselt[1]

[1] Institute of Human Communicative Research, & Division of Otorhinolaryngology, Department of Surgery, The Chinese University of Hong Kong
[2] Department of Electronic Engineering, The Chinese University of Hong Kong
[3] House Ear Institute, USA

ABSTRACT

Hearing aid amplification remains as the only effective intervention method for people suffering from sensory hearing loss with cochlea origin.

The design and validation of hearing aid amplification algorithms is based on the speech acoustics of non-tonal languages such as English. Chinese tonal languages (such as Cantonese and Mandarin) are fundamentally different from non-tonal languages that the meaning of a syllable changes with the perceived pitch configurations, with the phonetic structure of the syllable remains unchanged, thereby giving different lexical identities of the same syllable. These different pitch configurations are called lexical-tones which carry a significant amount of linguistic information in the speech signal. Given these fundamental acoustic-linguistic differences between tonal languages and non-tonal languages, the existing hearing aid signal processing algorithm developed for non-tonal languages may not give optimal performance for the hearing impaired population speaking Chinese tonal languages. The objective of this research was to identify which frequency region(s) has the temporal envelope(s) that carries important information for the identification of lexical-tones. Twenty normal-hearing native Cantonese speaking young adult subjects participated in the study. The results revealed that the lexical tone identification performance was the best when temporal envelopes from only the two high frequency bands (1-2kHz and 2-4kHz) were provided, whereas

the performance was the worst when temporal envelopes from only the two low frequency bands (60-500Hz and 500-1kHz) were provided. The findings suggested that high frequency bands are responsible for carrying important acoustic information for lexical-tone identification. The extracted temporal envelope carries more fundamental frequency harmonics from the male voice than from the female voice, which explains why lexical tone identification from the male voice was better than from the female voice. Further research is planned to investigate the effects of temporal envelope compression and expansion in different frequency regions on lexical-tone recognition and overall speech recognition in the normal-hearing and hearing-impaired populations.