<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Effects of excitation spread on the intelligibility of mandarin speech in cochlear implant simulations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author(s)</strong></td>
<td>Chen, F; Guan, T; Wong, LLN</td>
</tr>
<tr>
<td><strong>Citation</strong></td>
<td>The 8th International Symposium on Chinese Spoken Language Processing (ISCSLP 2012), Hong Kong, China, 5-8 December 2012. In International Symposium on Chinese Spoken Language Processing Proceedings, p. 35-39, article no. 6423502</td>
</tr>
<tr>
<td><strong>Issued Date</strong></td>
<td>2012</td>
</tr>
<tr>
<td><strong>URL</strong></td>
<td><a href="http://hdl.handle.net/10722/185165">http://hdl.handle.net/10722/185165</a></td>
</tr>
</tbody>
</table>
Effects of Excitation Spread on the Intelligibility of Mandarin Speech in Cochlear Implant Simulations

Fei Chen 1, Tian Guan 2, and Lena L. N. Wong 1

1 Division of Speech and Hearing Sciences, The University of Hong Kong
2 Graduate School at Shenzhen, Tsinghua University

Abstract

Noisy listening conditions remain challenging for most cochlear implant patients. The present study simulated the effects of decay rates of excitation spread in cochlear implants on the intelligibility of Mandarin speech in noise. Mandarin sentence and tone stimuli were processed by noise-vocoder, and presented to normal-hearing listeners for identification. The decay rates of excitation spread were simulated by varying the slopes of synthesis filters in noise-vocoder. Experimental results showed that significant benefit for Mandarin sentence recognition in noise was observed with narrower type of excitation. The performance of Mandarin tone identification was relatively robust to the influence of excitation spread. The results in the present study suggest that reducing the decay rates of excitation spread may potentially improve the speech perception in noise for cochlear implants in the future.

Index Terms – Cochlear implant, excitation spread, Mandarin speech perception

1. Introduction

Nowadays, cochlear implant (CI) is the only medical treatment to restore partial hearing to a severely-to-profoundly deafened person [1]. CI device inserts electrodes into the scala tympani of the cochlea, bypasses the hair cells, and directly stimulates the residual auditory nerves to elicit speech perception. As of 2008, over 120,000 patients globally have benefited from CI technique. Although many CI users are enjoying high levels of speech understanding in quiet, noisy listening conditions remain challenging for most. One limitation with cochlear implants is the difficulty stimulating spatially discrete spiral ganglion cell groups because of the channel interaction or the spread of current-excitation to generate the electrical field. A number of factors can influence the channel interaction in cochlear implants, e.g., the electrode configuration (e.g., bipolar or monopolar), and the distance between electrodes and spiral ganglion cells [2-3].

The effects of excitation spread on CI speech recognition remain unknown. Many studies hypothesized that bipolar electrode configuration (with an electrical-current decay rate of about 8 dB/mm, e.g., [4]) was more beneficial than monopolar configuration (with a decay rate of about 4 dB/mm) in terms of improving speech perception for CI patients. These hypotheses, however, need to be interpreted with caution given the large number of confounding factors associated with cochlear implants. It is not easy to assess or delineate the impact of each of those factors on speech perception due to interaction among them. Foremost among them is the variability in surviving nerve fiber sites and the differences in the distribution of neural sites containing excitable neuron elements among CI patients. These confounds make it extremely difficult to gauge the perceptual effect of excitation spread. For instance, subjects showed stronger preference for monopolar stimulation over bipolar stimulation in some studies [5].

Vocoder simulations have been used widely as an effective tool for assessing the effect of these confounding factors in the absence of patient-specific confounds (e.g., neuron survival sites) associated with cochlear implants [6-7]. These simulations have been shown by many to provide results consistent with the outcomes of CI patients. In these simulations, speech is processed in a manner similar to the CI speech processor and presented to normal-hearing (NH) listeners for identification (see review in [8]). Hence, vocoder simulations are used in this study to examine the effects of excitation spread on the intelligibility of Mandarin speech (sentences and tones). The listening experiments in the present study were designed to answer the following questions: (1) How would excitation spread affect the intelligibility of Mandarin speech in noise? (2) Does excitation spread have the same impact on the perception of Mandarin sentence as that of Mandarin tone?

2. Methodology

2.1. Subjects and stimuli

Seven (4 male and 3 female) NH native Mandarin-Chinese-speaking subjects participated in the experiment. The subjects’ ages ranged from 22 to 25 years, and the majority of subjects were graduate students at Tsinghua University. All subjects were paid for their participation.

The sentence material consisted of sentences taken from the Mandarin Hearing in Noise Test database [9]. There
were totally 12 lists in the MHINT corpus. Each MHINT list had 20 sentences, and each sentence contained 10 keywords. All the sentences were produced by a male speaker, with fundamental frequency ranging from 75 to 180 Hz. One adult male and one adult female native Mandarin-Chinese speaker produced the following six single-vowel syllables in each of the four Mandarin tones (/a/, /o/, /e/, /i/, /u/, /ü/) in a sound-treated booth, resulting in a total of 48 vowel tokens (=2 speakers × 6 vowels × 4 tones) for Mandarin tone identification. Both Mandarin sentences and vowels were recorded at a sampling rate of 16 kHz, and subsequently up-sampled to the rate of 44.1 kHz. A steady-state, speech-shaped noise (SSN), which has the same long-term average spectrum of testing material, was used to corrupt the target speech materials at −5, 0 and 5 dB signal-to-noise ratio (SNR) levels before vocoding processing.

2.2. Signal Processing

Stimuli were first processed through a pre-emphasis high-pass filter (2000 Hz cutoff) with a 3 dB/octave roll-off, and then band-passed into 22 frequency bands spanning the range of 180 to 7,938 Hz using sixth-order Butterworth filters. The Mel scale filter bank was used to allocate the 22 channels with the specified bandwidth. The envelope of the signal was extracted by full-wave rectification and low-pass filtering using a second-order Butterworth filter (400 Hz cutoff). The envelopes of each channel were modulated with white noise and re-filtered with the synthesis filters with various slopes.

The effect of excitation spread on speech intelligibility was simulated by varying the slopes of the synthesis bands in the noise vocoder. The extent of excitation spread along the basilar membrane was first calculated by dividing the electrical dynamic range (e.g., 15 dB) by the excitation decay rate (e.g., s=1, 4, and 16 dB/mm). The Greenwood frequency-to-place function was used to determine the range of frequency encompassing the spread of excitation [10]. The synthesized filters with specified magnitude were designed in Matlab using the fir2 function. Finally, the outputs of 13 channels with the largest root-mean-square energy of the envelopes amplitudes were summed to generate the synthesized stimuli. Figure 1 shows the block diagram of the noise-vocoder used in the present study, and Figure 2 exemplifies the magnitude responses of synthesis filters with slopes s=1, 4 and 16 dB/mm.

Bingabr et al. also assessed the effects of excitation spread on the intelligibility of English sentences and words in quiet and in noise [4]. The vocoder in this study is slightly different from that used by Bingabr et al. The band number...
Mean recognition scores for MHINT sentences and tones for all conditions are shown in Fig. 3. Statistical significance was determined by using the percent correct scores as the dependent variable, and SNR levels and decay rates of excitation spread as the two within-subject factors. For Mandarin sentence perception, two-way analysis of variance (ANOVA) with repeated measures indicated a significant effect ($F[2, 12]=82.79, p<0.0005$) of SNR level, a significant effect ($F[2, 12]=221.42, p<0.0005$) of decay rate of excitation spread, and a significant interaction ($F[4, 24]=7.87, p<0.0005$) between SNR level and decay rate of excitation spread. Post hoc tests, according to Fisher’s Least Significant Difference (LSD) test, revealed that the scores with $s=16$ dB/mm condition were significantly ($p<0.05$) higher than those obtained with $s=1$ dB/mm and $s=4$ dB/mm conditions at three SNR levels. Scores with $s=4$ dB/mm condition were significantly ($p<0.05$) higher than those obtained with $s=1$ dB/mm condition at three SNR levels. There was no significant difference between scores obtained with $s=16$ dB/mm and $s=4$ dB/mm conditions at 0 and 5 dB SNR levels.

For Mandarin tone identification, ANOVA with repeated measures indicated a significant effect ($F[2, 12]=20.62, p<0.0005$) of SNR level, a significant effect ($F[2, 12]=30.94, p<0.0005$) of decay rate of excitation spread, and a non-significant interaction ($F[4, 24]=1.87, p=0.1297$) between SNR level and decay rate of excitation spread. Post hoc tests only revealed that the scores with $s=4$ dB/mm condition were significantly ($p<0.05$) higher than those obtained with $s=1$ dB/mm condition at three SNR levels. There was no significant difference between scores obtained with $s=16$ dB/mm and $s=4$ dB/mm conditions at 0 and 5 dB SNR levels.
4. DISCUSSIONS AND CONCLUSION

Bingabr et al. showed that, in noise-vocoder based simulation study, reducing the spread of excitation had little effect on recognition of English sentences and words in quiet, when the spectral resolution was adequate (i.e., >8 channels); however, a comparatively larger effect was observed with recognition of sentences in noise [4]. The present studies also suggested the potential advantage of using more focused excitation to stimulate auditory nerves and elicit hearing, especially in noise. Recently, novel technologies are being developed to use more focused excitation to stimulate auditory nerves, e.g., optical stimulation [11]. However, besides the decay rate of excitation spread, numerous factors (e.g., duration of deafness, and surviving neural pattern) may also affect the performance of CI users in quiet and noisy conditions; hence, further studies need to be performed to manifest the advantages of the focused excitation in cochlear implants.

The present results show that, compared with the more pronounced response to noise level and decay rate of excitation observed in Mandarin sentence recognition task, the performance of Mandarin tone identification is relatively robust. This indicates that factors other than tone identification account more for the degradation of Mandarin sentence understanding. For instance, the acoustic boundaries or landmark could play a more important role in speech perception in noise [12].

The findings in this study also shed light on investigating the relationship between Mandarin tone identification and sentence recognition in maskers. Although many studies have examined Mandarin tone identification in quiet [13], only a limited number of studies have investigated how CI hearing performs in tone-identification task with noise being present. The robustness of Mandarin tone recognition in noise might be partially due to the following two factors. First, acoustic cues on lexical tone identification mainly exist in vowel segments, and thus might be better recognized, as vowel segments usually have stronger energies and more favorable local SNRs than other segments (e.g., consonants). Second, there are only four choices for tones, and such a higher chance level may make tone recognition more robust than consonant or vowel recognition. More work still needs to be done to examine the relationship between Mandarin sentence recognition and tone identification, particularly in noise.

In conclusion, the present study evaluated the effects of excitation spread on the intelligibility of Mandarin speech in cochlear implant simulations. The decay rates of excitation spread were simulated by varying the slopes of synthesis filters in noise vocoder. Experimental results showed that significant benefit for Mandarin sentence recognition in noise can be obtained with narrower type of excitation. In addition, the performance of Mandarin tone identification was relatively resistant to the influence of excitation spread, in contrast to that of Mandarin sentence recognition.

5. ACKNOWLEDGEMENT

This research was supported by Faculty Research Fund, Faculty of Education, The University of Hong Kong, and by General Research Fund (GRF), administered by the Hong Kong Research Grants council. This work was also supported Grant 30800234 and 31271056 from National Natural Science Foundation of China, and Grant 10151805702000000 from Guangdong Natural Science Foundation. The authors would like to thank Ms. Hong Sun for her assistance in data collection.

6. REFERENCES
