EFFECT OF FAST AGC ON COCHLEAR IMPLANT SPEECH INTELLIGIBILITY

Phyu P. Khing1,2, Eliathamby Ambikairajah1, Brett A. Swanson2

1 School of Electrical Engineering and Telecommunications
The University of New South Wales, Sydney, NSW 2052, Australia
2 Cochlear Ltd, Sydney, NSW 2066, Australia

ABSTRACT

This study investigated the effect of fast-acting Automatic Gain Control (AGC) on the speech intelligibility of cochlear implant users as a function of presentation level. Both low and high signal-to-noise ratio conditions were investigated. The AGC substantially reduced the amount of clipping, but did not give consistent improvements in intelligibility. With no AGC, and high signal-to-noise ratio, speech scores were not significantly degraded until more than 25% of stimulation pulses were affected by clipping.

Index Terms — cochlear implant, speech intelligibility, Automatic Gain Control, compression

1. INTRODUCTION

The dynamic range of electrical current pulses in a cochlear implant (5 – 20 dB) is very small compared to the dynamic range of acoustic signals encountered in everyday life [1]. Thus an ongoing challenge has been how to best convey the information in the acoustic signals onto the electrodes. The overall level of everyday speech may vary in a 35 dB range from casual conversation to shouting [2]. The individual components of a segment of speech vary over a 40 – 50 dB range even when the average presentation level is fixed [3]. Thus all cochlear implant systems incorporate some form of compression to accommodate a wide range of acoustic signals. Amplitude levels beyond a specified maximum level will be clipped to the maximum level during mapping. This causes amplitude distortion. One goal of this study was to determine the effect of this clipping on the speech intelligibility performance of CI recipients.

There are a number of studies in the literature that observed the word recognition performance of the normal hearing and hearing impaired subjects at higher-than-normal speech and noise levels. In the study done by Studebaker et al., they observed a decline in word recognition performance of the normal hearing and hearing impaired listeners when the speech level was increased above the normal conversation level in a fixed SNR condition [4]. The performance degradation was significantly higher in lower SNR conditions. The same observation was made by Dubno et al. in [5]. Both studies indicated the relative decrease in the effective signal-to-noise ratio with the increase in speech level as the main cause of performance degradation. Although we have not encountered a similar study with the CI recipients, a few studies observed the effects of the modification in amplitude cues to the speech intelligibility performance of CI users. In the study done by Zeng and Galvin, they examined the effect of electric dynamic range reduction on the speech recognition performance of the CI recipients using the Nucleus-22 SPEAK processor [6]. They modified the processing so that the amplitude levels of acoustic signals were represented by a binary form in electrical current domain. According to their study, the speech intelligibility in quiet conditions was not significantly affected by the maximum reduction in the electric dynamic range. Therefore, the retention of spectral envelope shape is not essential to speech recognition of CI listeners using spectral envelope processing strategies in quiet conditions. In the study done by Zeng et al. with ten CI subjects, they observed a significant performance reduction with the decrease in the input dynamic range (IDR) below 40 dB in quiet condition [3]. These two studies indicate that the limitation imposed on the acoustic amplitudes before mapping affects the speech intelligibility more than the limitation imposed on the electrical current amplitudes after mapping.

Based on the findings from these studies, we hypothesized that the speech intelligibility of CI recipients would degrade significantly at higher-than-normal presentation levels when the clipping distortion started to reduce the effective SNR level. The rate of degradation was expected to be faster in lower SNR conditions. If we can correlate the amount of clipping distortion to the speech scores, it can serve as a control parameter in an intelligent gain algorithm before amplitude mapping to improve speech performance regardless of presentation level.

A fast AGC is part of the everyday sound processing strategy of the Nucleus systems. The main aim of fast compression in cochlear implants is to prevent discomfort due to excessive stimulation at maximum comfortable level (C-level) by compressing the incoming acoustic signals above the knee level at the frontend. It therefore helps to control loudness. Thus the performance of the CI listeners
with AGC is expected to be better in higher-than-normal presentation levels than without AGC. Another goal of this study was to observe how much fast AGC helped to reduce the proportion of current pulses hitting C-level and the relative improvement in speech intelligibility by the amount of reduction. There are many studies in the literature that investigated side effects of fast AGC. Stone and Moore examined the negative effects of fast compression on normal hearing subjects using cochlear implant simulator [7]. They attributed the comodulation between the target speech and background noise introduced by the wideband fast compression as the main cause of performance degradation in their experiment. In the later study done by the same authors, they quantified the effects of fast AGC as three types of envelope change in the target speech when compression was applied to the mixture of the target speech and masker speech [8]. All three measures indicated the fast compression had negative effects on the envelope of the target speech. They pointed out that the partial correlation between the target and masker introduced by the fast AGC, also known as comodulation, was the dominant factor affecting intelligibility.

2. METHOD

Four post-lingually deafened adult cochlear implant recipients, mean age 75, participated in this study. The subjects were unilaterally implanted with either the Nucleus 24 or Nucleus Freedom cochlear implant. All had more than two years experience with their implant, and previous experience with speech tests in clinical evaluations.

Listening tests were carried out in a sound treated room. Sentences and noise were presented from a single loudspeaker approximately 1 metre in front of the subject. The sentences were presented in a background of 4-talker babble noise. The presentation level was varied from 55 to 89 dB SPL. For each target presentation level, the background noise level was set to give a fixed SNR. Both low and high SNR conditions were tested. The high SNR was 20 dB for all subjects. The low SNR was set to 10 dB for three subjects (S02, S03, S04). S01 scored too well at 10 dB SNR, so the low SNR condition was reduced to 6.5 dB for S01. Two signal processing conditions were tested: AGC on and AGC off.

Sentences were scored by the number of morphemes correct. Eight sentences were presented for each presentation level, and each sentence contained an average of 6 morphemes. Statistical analysis was performed using Monte Carlo simulation, assuming binomial distributions [9].

All subjects used the Advanced Combination Encoder (ACE) strategy (Figure 1). Subjects’ processing parameters are summarized in Table 1.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Implant</th>
<th>Number of channels</th>
<th>Number of maxima</th>
<th>Stimulation rate (pps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>S01</td>
<td>CI24RE</td>
<td>20</td>
<td>10</td>
<td>1800</td>
</tr>
<tr>
<td>S02</td>
<td>CI24M</td>
<td>22</td>
<td>12</td>
<td>720</td>
</tr>
<tr>
<td>S03</td>
<td>CI24RE</td>
<td>22</td>
<td>8</td>
<td>900</td>
</tr>
<tr>
<td>S04</td>
<td>CI24RE</td>
<td>22</td>
<td>8</td>
<td>900</td>
</tr>
</tbody>
</table>

Table 1. Subject stimulation parameters

To determine the amount of clipping, one sentence was selected and played at each presentation level and AGC condition for SNRs of 10 and 20 dB. The resulting stimulus pulses were captured using an engineering tool, the Decoder Implant Emulation Tool (DIET). The captured pulse data was then analysed to determine the proportion of pulses that had been delivered at C-level.
3. RESULTS

The top panel of Figure 2 shows the mean scores averaged across subjects. The bottom four panels of Figure 2 show the scores of each individual subject. The clear trend for all conditions was that mean scores decreased with increasing presentation level. Figure 3 shows the proportion of current pulses hitting C-level.

3.1. Without AGC

Without AGC, Figure 3 shows that the amount of clipping grows monotonically as the presentation level exceeds 65 dB SPL. At high SNR, mean scores across subjects were consistently between 85% and 95% up to a presentation level of 83 dB SPL; at this level, approximately 25% of pulses were being clipped to C-level. Subjects differed in their tolerance to this clipping. Subjects S01 and S02 scored consistently well and showed little degradation in their scores for high presentation levels, while S03 and S04 were more sensitive to clipping. The paired test showed no significance for S01, marginal significance for S02 (p = 0.04) and high significance (p < 0.001) for S03 and S04 when comparing the scores above and below 83 dB SPL.

At low SNR, mean scores were lower (70 – 80%), and began to fall off beyond a presentation level of 70 dB SPL; at this level, only 9% of pulses were being clipped. The fall-off started as early as 65 dB SPL for S04. According to individual subject results, the scores above and below the presentation level of 70 dB SPL were statistically significant (p < 0.001) for S02, S03 and S04. The fall-off trend of S01 was different from the other subjects. S01 was more tolerant of noise at low SNR than the other subjects.

3.2. With AGC

As expected, AGC had no effect on mean scores for presentation levels below 65 dB SPL, as the peaks of the speech at this level just reached the AGC knee-point. For presentation levels above 70 dB SPL, the AGC gave a small improvement in mean scores at low SNR, and to a lesser extent at high SNR. Examining individual subject’s results, a significant improvement in scores at low SNR was obtained by subjects S02 (p = 0.01), S03 (p < 0.001), and S04 (p = 0.01). At high SNR, the only subjects to obtain a significant benefit were S02 (p = 0.003) and S04 (p < 0.001); while S01 suffered a significant (p = 0.002) drop in scores with AGC. The fall-off trends of scores with and without AGC were very similar regardless of SNR.

Figure 2. Mean and individual scores of the four subjects

Figure 3. The proportion of the pulses hitting C-level
4. DISCUSSION

Referring to Figure 3, AGC was successful in reducing the amount of clipping (by more than half) in both SNRs. Despite this, AGC did not change the overall trend in the scores, with scores still degrading as presentation level increased. This supports the hypothesis that clipping was not the major cause of performance degradation. It appeared that the positive effects of AGC on clipping were offset by its negative effects on the envelope of target speech at high presentation levels in both SNRs. AGC reduced the effective SNR by compressing the high level components of the speech more than the background noise components in high SNR. In low SNR, AGC compressed both speech and noise components, introducing partial correlation between them. As found by Stone and Moore, a wide band fast compression comodulated speech and noise together in a way that they became perceptually inseparable [8]. Without AGC, both clipping distortion and the reduction in effective SNR caused the performance degradation. It appeared that the participants were more tolerant to the amplitude distortion due to clipping than the background noise. As presentation level increases, clipping initially affects only the target speech, reducing its effective level, while having little impact on the noise level. Thus clipping reduces the effective SNR. For both SNR conditions, the scores began to fall off when the noise level was relatively high, about 60 dB SPL. The effective SNR at this point was considerably less than the input SNR level. This agrees with previous research that noise is more detrimental than amplitude distortion to speech intelligibility [6].

Since the CI subjects were able to tolerate the amplitude distortion due to clipping, the use of the fast AGC at high levels to reduce the excessive C level stimulation was not as effective as we expected. Although, we observed some increase in scores with fast AGC at high presentation levels in both SNRs, it was not consistent across the subjects. The overall trend of scores with AGC was still similar to no AGC in both SNR conditions. It showed that fast AGC offered little help for speech in various presentation and background levels.

We attributed the effective SNR drop with the increase in presentation level as the main cause of performance degradation as in the case of the normal hearing listeners [4, 5]. For normal hearing subjects the relative increase in masking level at high presentation levels caused the drop in effective SNR level. For CI subjects, the compression to adjust the mismatch between the acoustic and electric dynamic range together with the relative increase in masking level caused the effective SNR drop at high presentation levels. The degree of performance degradation is far worse for the CI recipients since they rely mainly on the envelope information of target speech which is highly sensitive to the competing noise in the background.

5. CONCLUSION

In the absence of the usual slow-acting compression, the speech intelligibility of CI recipients degraded when listening to speech in higher-than-normal presentation levels. The performance degradation was mainly due the reduction in effective SNR. The effect started to be significant when the noise level was more than 60 dB SPL. At this level, the current pulses corresponding to the noise started to produce C-level stimulation. The fast AGC helped to reduce the proportion of stimuli hitting C level, yet it only provided small improvements in intelligibility. The mean performance at high SNR condition was robust even with 25% of the pulses hitting C-level. This implies that if the usual slow-acting compression was available to slowly adjust the overall level, then short intervals of clipping due to louder transients would not be very objectionable, and there would be little need for fast AGC.

6. REFERENCES