PREDICTING THE EFFECT OF AGC ON SPEECH INTELLIGIBILITY OF COCHLEAR IMPLANT RECIPIENTS IN NOISE

Phyu P. Khing[1,2], *Eliathamby Ambikairajah*[1], *Brett A. Swanson*[2]

[1] School of Electrical Engineering and Telecommunications
The University of New South Wales, Sydney, NSW 2052, Australia
[2] Cochlear Ltd, Sydney, NSW 2109, Australia

ABSTRACT

The aim of this study was to predict the effects of automatic gain control (AGC) on the speech intelligibility of cochlear implant recipients in noise. Two simple signal metrics were calculated: the proportion of clipping, and the output signal-to-noise ratio (SNR). Psychometric functions were fitted to the percent correct scores averaged over five cochlear implant recipients for three different AGC conditions, at two different input SNRs, for a range of presentation levels. The output SNR was a good predictor of the recipients' mean scores.

Index Terms— cochlear implant, AGC, speech intelligibility, SNR, psychometric fit

1. INTRODUCTION

An ongoing challenge in cochlear implant (CI) sound processing is how to best convey the information in acoustic signals onto the electrodes, because the dynamic range of electrical current pulses is very small (5 – 20 dB) compared to the dynamic range of acoustic signals (~120 dB). The overall level of speech varies in a 30 – 40 dB range from casual conversation to shouting [1, 2]. For a fixed presentation level, the level of short speech segments varies over a 40 – 50 dB range [3].

When a CI recipient is fitted, the threshold level (T-level) and maximum comfortable level (C-level) are determined for each electrode. To avoid excessive loudness, the stimulation current is not allowed to exceed the C-level. In the Nucleus CI system, the automatic gain control (AGC) handles variation in the overall speech level. It is followed by the Loudness Growth Function (LGF), an instantaneous non-linear compression, which compresses a (typically) 40 dB dynamic range into the available current range. The shape of the LGF is intended to make the CI recipient's loudness perception match that of a normal hearing person for changes in sound intensity. The LGF saturation level is the input level that produces current at C-level. If the input level exceeds the saturation level then clipping occurs.

It is well known that low rate envelope modulation is very important for speech intelligibility [4, 5]. Moreover speech envelope peaks are perceptually most relevant for intelligibility [6]. Peak clipping at the LGF reduces the modulation depth of the envelopes, and flattens the short-term spectral profile. As a result, the speech intelligibility of CI recipients may be degraded. Thus the purpose of the AGC is to avoid clipping at the LGF. An AGC can be fast-acting to adjust the intensity of individual component of speech or slow-acting to adjust the overall level of word or sentences.

Khing et al. studied the speech intelligibility of CI users with and without the frontend fast-acting AGC while other slow gain algorithms were switched off [7]. A significant proportion of clipping occurred at high presentation levels when no AGC was used. The frontend AGC reduced the amount of clipping by half, yet the score improvement was modest. The effects of a fast-acting compression on the intelligibility of speech in noise were analyzed by Stone and Moore [8-10]. The most important factor that degraded the speech intelligibility was the common modulation, introduced by the fast compression, between the target speech and the interference. They proposed the ASMC metric to quantify this effect [10]. Furthermore, fast-acting compression can reduce the low rate modulation of speech and introduce other types of nonlinear distortion and therefore degrade speech intelligibility [11].

Speech intelligibility tests are time-consuming for both CI recipients and researchers. A signal metric that reliably predicted CI recipients' speech intelligibility would accelerate development and optimization of CI processing algorithms. A number of metrics have been proposed, such as the speech intelligibility index (ANSI S3.5-1997) [12] and the speech transmission index [13, 14]. Chen and Loizou evaluated the performance of a number of metrics in predicting the speech intelligibility of vocoded speech [15]. They found that the coherence-based measures and the STI-based measures had a high correlation with the intelligibility scores of normal hearing subjects on vocoded speech in noise.

Sound processing, either linear or nonlinear, may cause the SNR of the output signal to be different from the SNR of the input stimuli and therefore affect the intelligibility [16, 17]. The nonlinear operations of the CI signal path are maxima selection and the LGF (Figure 1). The two signal

978-1-4799-0356-6/13/$31.00 ©2013 IEEE

metrics investigated in the present study are the clipping proportion and the output SNR. The clipping proportion measures how often peak clipping occurs, and is a measure of the effectiveness of the AGC system, but may not be directly related to speech intelligibility. Rhebergen et al. developed the apparent SNR method and investigated wide dynamic range compression [18]. The present study extended the apparent SNR method to apply to CI systems. The goal was to determine whether these signal metrics could predict the speech-in-noise scores of CI recipients under a variety of processing and stimulus conditions.

2. METHOD

Sentences were presented to each CI recipient in a background of 4-talker babble noise, from a single loudspeaker in a sound-treated room. The noise started one second before each sentence. Each sentence list had a fixed presentation level and a fixed input SNR. Lists were presented at levels ranging from 55 to 89 dB SPL, at two input SNRs, 10 and 20 dB. The recipient verbally repeated each sentence, and was scored on the number of morphemes (word parts) correct. The data set comprised the mean percent correct scores of five CI recipients using three different AGC processing conditions. Results for two of these conditions were reported previously [7].

2.1. CI Sound Processing

The signal path is shown in Figure 1. The ADC sampled the signal at 16 kHz. The filter bank divided the signal into 22 bands, and was followed by quadrature envelope detection. The Maxima Selection block examined the envelopes in each analysis period and selected those with the largest amplitude for stimulation. The LGF applied instantaneous non-linear compression. The Amplitude Mapping block produced stimulation pulses with current levels ranging from T-level to C-level for each electrode. One or both AGCs were disabled to give three conditions: (1) no AGC, (2) frontend AGC, and (3) multichannel AGC. The usual slow-acting AGC stages (ASC and ADRO) were disabled for the purposes of this study.

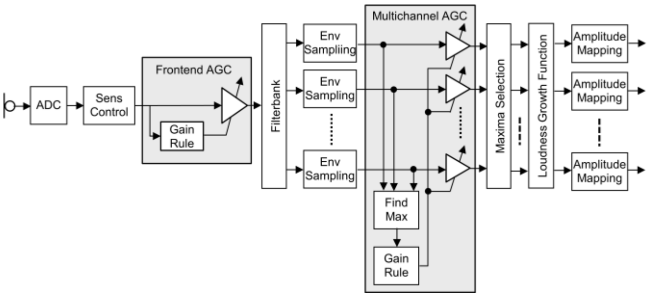

Figure 1 CI sound processing with the frontend and multichannel AGC

2.1.1. Frontend AGC

The frontend AGC was a single-channel fast-acting AGC located before the filter bank (the standard algorithm for the Nucleus Freedom sound processor). It reduced the gain when the envelope of the input waveform exceeded the compression threshold. The attack time was 5 ms, the release time was 75 ms, the compression threshold for a 1 kHz tone was 73 dBSPL, and it had an infinite compression ratio.

2.1.2. Multichannel AGC

The multichannel AGC was a new gain algorithm developed in the present study. It was located after the filter bank. It reduced the gain when the largest of the filter bank envelopes exceeded the compression threshold. The same gain was applied to all channels. The compression threshold was set equal to the saturation level of the LGF, ensuring that no clipping occurred at the LGF. It had zero attack time, the release time was 625 ms, the compression threshold for a 1 kHz tone was 59 dBSPL, and it had an infinite compression ratio. The apparent difference in compression thresholds of the two AGCs was due to the use of a 1 kHz tone for calibration. For a speech signal, both AGCs began to reduce the gain at a level of 65 dBSPL.

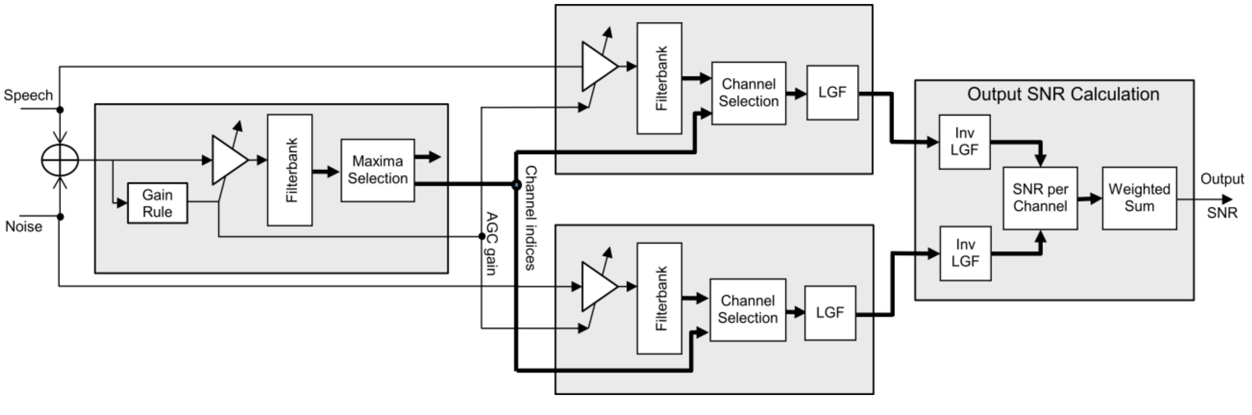

Figure 2 Output SNR calculation for cochlear implant ACE signal path. For the illustration purpose, only the frontend AGC is included in the signal path in this diagram. Similar setup is used for the multichannel AGC where the gain is taken after the filterbank.

2.2. Signal metrics

The processing and input conditions that were tested with the CI recipients were replicated in MATLAB. Five sentences were concatenated, with silent gaps of 4 seconds between sentences, so that each AGC would start with 0 dB gain for each sentence. Both metrics were calculated over the duration of the sentences, excluding the silent gaps. A psychometric function was fitted to the recipient mean percent correct scores for each signal metric using the psignifit toolbox for MATLAB by Jeremy Hill (version 2.5.6, available at http://bootstrap-software.org/psignifit/), which implements a maximum-likelihood method [19]. The goodness of fit was quantified by the deviance, D; a smaller deviance indicated a better fit [19].

2.2.1. Clipping Proportion

The clipping proportion was calculated by counting the number of samples at the input to the LGF that exceeded the saturation level, and dividing by the total number of samples. By design, the clipping proportion was zero for the multichannel AGC.

2.2.2. Output SNR

The output SNR calculation was similar to Rhebergen's apparent SNR calculation [17, 18], but adapted for CI processing, as shown in Figure 2. Firstly, the speech and noise mixture was processed through the signal path, and two resulting signals were recorded: the gain applied by the AGC (if present), and the channel indices from the maxima selection block. Next, the clean speech was processed through the signal path, applying the recorded gain, and choosing stimulation pulses using the recorded channel indices. An inverse LGF was applied to revert to the linear domain, while retaining the effect of clipping. Similarly, the noise alone was processed, using the recorded gain and channel indices. Then, the SNR was calculated for each channel. Finally, the channel SNRs were weighted according to their relative signal power, as in [20], and summed to give the output SNR. An alternative method was also investigated, setting the channel weights according to the articulation index band importance function from ANSI S3.5 1997, but it made little difference to the metric.

3. RESULTS

Figure 3 shows the mean percent correct scores of 5 CI participants for a total of 6 conditions (i.e., 3 AGC conditions x 2 input SNR conditions) as a function of presentation level. In each condition, scores degraded as presentation level increased, and were worse for the lower input SNR. Scores were lowest with no AGC, better with the frontend AGC, and best with the multichannel AGC.

Figure 4 is a scatter plot of the scores against the clipping proportion for the no-AGC and frontend AGC conditions (the multichannel AGC condition is not shown because the clipping proportion was zero). The bottom panel shows the 10 dB SNR scores, the middle panel shows the 20 dB SNR scores, and the top panel pools both sets of scores. Good psychometric fits were obtained for each condition (bottom and middle panels), implying that clipping has a detrimental effect on scores. However, the psychometric functions have very different shapes for the no-AGC and frontend AGC conditions, and the top panel shows that fitting one psychometric function to the pooled scores yielded a very poor fit (high deviance). Compared to no AGC, the frontend AGC substantially reduced the clipping proportion (to less than 20%), but the scores degraded more rapidly as a function of clipping proportion. Furthermore, for the same amount of clipping, scores were higher at the better input SNR. Thus clipping proportion alone was not a good predictor of scores, and some other factor influenced the scores.

Figure 5 is a scatter plot of the scores against the output SNR. The psychometric functions at 20 dB SNR (middle panel) are very similar for all three AGC conditions. At 10 dB SNR (bottom panel), there are some differences: the

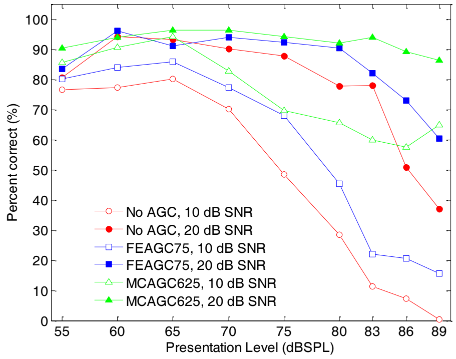

Figure 3 Mean percent correct scores of 5 CI participants

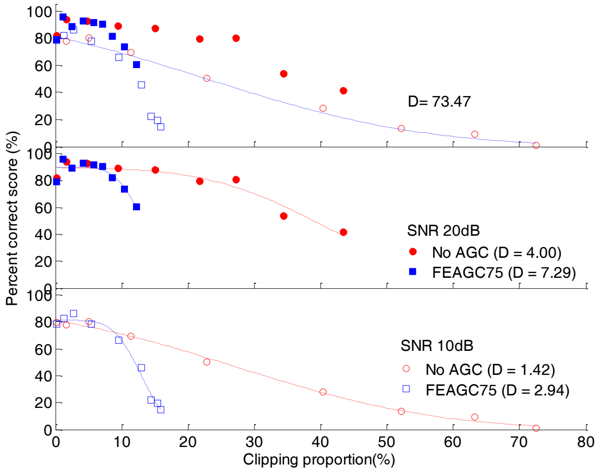

Figure 4 The percent correct scores of 5 CI subjects as a function of the clipping proportion

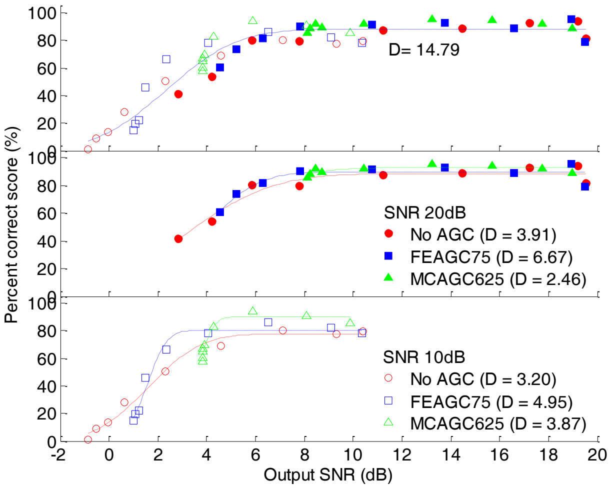

Figure 5 The percent correct scores of 5 CI subjects as a function of the output SNR

multichannel AGC had higher asymptotic scores, and scores rolled off at a higher output SNR. However, the top panel shows that one psychometric function can provide a relatively good fit to the pooled scores. Thus output SNR was a good predictor of scores for all three AGC conditions, at both input SNRs, across a wide range of presentation levels.

4. DISCUSSION

CI recipients rely on envelope cues for speech perception, so it was expected that clipping, a form of envelope distortion, would have a detrimental effect on speech intelligibility. However, clipping proportion by itself was not a good predictor of scores; and eliminating clipping (with the multichannel AGC) does not provide perfect performance. Performance was clearly affected by noise, which the clipping proportion metric does not capture.

Because CI sound processing is non-linear, the SNR at the output differs from the SNR at the input. The present study extended prior work on calculating the output SNR, to make it suitable for CI processing. Prior work recorded the gains produced by the AGC in response to the speech and noise mixture, and applied these gains separately to the clean speech and the noise [17]. However, if the remaining CI processing was then applied, the maxima selection on the clean speech would choose different channels from the maxima selection on the noise. Instead, the channel indexes were recorded from the maxima selection on the speech and noise mixture, so that the contributions of speech and noise to each stimulation pulse could be identified.

Output SNR was a better predictor of the scores than clipping proportion, which implies that noise had a larger impact on speech intelligibility than envelope distortion in these listening tests.

These results suggest areas for future research. The output SNR metric could be compared to the ASMC metric [10]. For the same input SNR, CI recipients perform better with stationary noise than with the 4-talker babble used in the present study [21]. A good metric should be able to predict this.

The output SNR metric is unable to predict intelligibility for speech in quiet. At high presentation levels, clipping will reduce intelligibility. Metrics such as STI-based measures can quantify the temporal modulation reduction, and the articulation index and coherence-based metrics can quantify the spectral envelope distortion. It may be possible to combine these metrics with the output SNR metric.

Finally, a metric such as output SNR can be used to guide the selection of processing parameters, such as AGC release time, or in the development of new algorithms.

5. CONCLUSION

Although clipping is detrimental, the clipping proportion metric alone was not a good predictor of the speech intelligibility of CI recipients. The novel output SNR metric developed in the present study is simple, easy to implement, and was an effective predictor of CI speech intelligibility scores for a range of processing, presentation level, and input SNR conditions.

6. REFERENCES

[1] K. S. Pearsons, R. L. Bennett and S. Fidell, "Speech levels in various noise environments," Office of Health and Ecological Effects, Office of Research and Development, US EPA, Washington, DC1977.

[2] W. O. Olsen, "Average Speech Levels and Spectra in Various Speaking/Listening Conditions: A Summary of the Pearson, Bennett, & Fidell (1977) Report," *Am J Audiol,* vol. 7, pp. 21-25, October 1, 1998 1998.

[3] F. Zeng*, et al.*, "Speech dynamic range and its effect on cochlear implant performance," *The Journal of the Acoustical Society of America,* vol. 111, p. 377, 2002.

[4] R. Drullman, J. M. Festen and R. Plomp, "Effect of reducing slow temporal modulations on speech reception," *Journal of the Acoustical Society of America,* vol. 95, pp. 2670-2680, 1994.

[5] C. Füllgrabe, M. A. Stone and B. C. J. Moore, "Contribution of very low amplitude-modulation rates to intelligibility in a competing-speech task," *The Journal of the Acoustical Society of America,* vol. 125, p. 1277, 2009.

[6] R. Drullman, "Temporal envelope and fine structure cues for speech intelligibility," *Journal of the Acoustical Society of America,* vol. 97, pp. 585-592, 1995.

[7] P. P. Khing, E. Ambikairajah and B. A. Swanson, "Effect of fast AGC on cochlear implant speech intelligibility," in *Acoustics,*

Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on, 2011, pp. 285-288.

[8] M. A. Stone and B. C. J. Moore, "Effect of the speed of a single-channel dynamic range compressor on intelligibility in a competing speech task," *The Journal of the Acoustical Society of America,* vol. 114, p. 1023, 2003.

[9] M. A. Stone and B. C. J. Moore, "Side effects of fast-acting dynamic range compression that affect intelligibility in a competing speech task," *The Journal of the Acoustical Society of America,* vol. 116, p. 2311, 2004.

[10] M. A. Stone and B. C. J. Moore, "Quantifying the effects of fast-acting compression on the envelope of speech," *The Journal of the Acoustical Society of America,* vol. 121, p. 1654, 2007.

[11] R. Plomp, "Noise, amplification, and compression: Considerations of three main issues in hearing aid design," *Ear and hearing,* vol. 15, p. 2, 1994.

[12] A. ANSI, "S3. 5-1997, Methods for the calculation of the speech intelligibility index," *New York: American National Standards Institute,* 1997.

[13] H. J. M. Steeneken and T. Houtgast, "A physical method for measuring speech transmission quality," *The Journal of the Acoustical Society of America,* vol. 67, p. 318, 1980.

[14] T. Houtgast and H. J. M. Steeneken, "A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria," *The Journal of the Acoustical Society of America,* vol. 77, p. 1069, 1985.

[15] F. Chen and P. C. Loizou, "Predicting the intelligibility of vocoded speech," *Ear and hearing,* vol. 32, p. 331, 2011.

[16] B. Hagerman and A. Olofsson, "A method to measure the effect of noise reduction algorithms using simultaneous speech and noise," *Acta Acustica united with Acustica,* vol. 90, pp. 356-361, 2004.

[17] K. Rhebergen, N. Versfeld and W. Dreschler, "The dynamic range of speech, compression, and its effect on the speech reception threshold in stationary and interrupted noise," *The Journal of the Acoustical Society of America,* vol. 126, p. 3236, 2009.

[18] K. S. Rhebergen, N. J. Versfeld and W. A. Dreschler, "Quantifying and modeling the acoustic effects of compression on speech in noise," *The Journal of the Acoustical Society of America,* vol. 123, pp. 3167-3167, 2008.

[19] F. A. Wichmann and N. J. Hill, "The psychometric function: I. Fitting, sampling, and goodness of fit," *Percept Psychophys,* vol. 63, pp. 1293-313, Nov 2001.

[20] J. Ma, Y. Hu and P. C. Loizou, "Objective measures for predicting speech intelligibility in noisy conditions based on new band-importance functions," *The Journal of the Acoustical Society of America,* vol. 125, p. 3387, 2009.

[21] G. S. Stickney*, et al.*, "Cochlear implant speech recognition with speech maskers," *The Journal of the Acoustical Society of America,* vol. 116, p. 1081, 2004.