Effect of Adaptive Envelope Compression in Simulated Electric Hearing in Reverberation

Ying-hui Lai
Research Center for Information Technology Innovation, Academia Sinica, Taipei, Taiwan
jackylai@citi.sinica.edu.tw

Fei Chen
Division of Speech and Hearing Sciences, The University of Hong Kong, Hong Kong, China
feichen1@hku.hk

Yu Tsao
Research Center for Information Technology Innovation, Academia Sinica, Taipei, Taiwan
yu.tsao@citi.sinica.edu.tw

***Abstract*–The narrow dynamic range for speech perception partially accounts for the poor speech understanding abilities of hearing-impaired patients fitted with cochlear implants, particularly in challenging listening conditions, e.g., in reverberation. Wide dynamic range compression is designed to compress speech signal into the usable hearing dynamic range of implanted patients; however, it normally uses a static compression based strategy. An adaptive envelope compression (AEC) strategy was recently proposed for speech processing in cochlear implants. It implemented the envelope compression as close to linear as possible, while confined the compressed amplitude envelope within the pre-set dynamic range. This study further assessed its effect to improve speech perception in reverberation. Vocoder simulation experiment showed that, when narrowed down to a small dynamic range, the AEC-processed sentences could yield a higher intelligibility score in reverberation than the static compression processed sentences.**

***Index Terms*–cochlear implants, dynamic range, adaptive envelope compression, vocoder simulation**

1. INTRODUCTION

Cochlear implant (CI) is presently the only electronic device that could provide a sense of sound to a person with profound-to-severe hearing loss [1-2]. In a CI device, the input speech signal is received by an external microphone and fed into a speech processor. The speech processor normally captures the multi-channel temporal envelopes from the input speech signal, and then generates electric stimulations to directly excite the residual auditory nerves [1-2]. Due to biological constraints, the dynamic range (DR) of stimulation generated by a speech processor in CI is much smaller than that of the real speech signal (e.g., 5 dB vs. 50 dB). Hence, a compression scheme is required to compress the DR of the input signal to a desirable level. Previous studies indicated that the dynamic range of temporal envelope was an important factor predicting the speech intelligibility for CI users [3-5].

Most of the present CI devices use fixed compression function in converting acoustic amplitude envelope to electric current signal. While this fixed mapping function confines the overall electric current within a small dynamic range, this strategy is not optimized in terms of making best use of the small hearing dynamic range for speech perception. Hence, several adaptive dynamic range control strategies have been developed [e.g., 6-9]. Lai et al. recently proposed an adaptive envelope compression (AEC) strategy to confine the amplitude envelope of speech signal within a fixed dynamic range, but continuously adjust compression ratio for short-term amplitude [7]. Through this processing, the local dynamic range approached to that of the uncompressed amplitude envelope, and the amplitude envelope processed by the AEC strategy led to a higher intelligibility for noise-corrupted sentences than that processed by the static envelope compression (SEC) strategy [7].

In addition to noise interference, reverberation is well known to change speech quality and can also impact the speech intelligibility of implanted patients, as it blurs temporal and spectral cues and flattens formant transitions [e.g., 10]. Following the previous study on AEC processing for speech perception in noise [7], the purpose of this study is to assess the effect of AEC processing to improve speech understanding in reverberation for cochlear implants. More specifically, this study will compare the recognition of speech synthesized by the AEC and SEC strategies. To assess the effect of dynamic range compression to speech perception in reverberation, vocoder simulation involving normal-hearing (NH) subjects will be used in this study. Vocoder simulation has been extensively used to study the effects of various factors on speech perception by CIs, because of its advantages of avoiding the impact of patient-specific confounding factors (e.g., neural surviving pattern) existing in clinical populations [e.g., 1, 2, 11].

2. METHODS

2.1. Subjects and materials

Nine (19-27 yrs., 4 female) NH native-Mandarin speakers participated in the listening experiment. Sentence lists from the Mandarin version of Hearing in Noise Test (MHINT) were used as the testing materials in this study [12]. All sentences were pronounced by a male native-Mandarin speaker with fundamental frequency ranging from 75 to 180 Hz, and recorded with a 16 kHz sampling rate. Head related transfer functions recorded in a 5.5 m × 4.5 m × 3.1 m (length × width × height) room with a total volume of 76.8 m^3 [13] were used to simulate the reverberant conditions. The initial average reverberation time of the experimental room (T_{60} = 1.0 s) was reduced to T_{60} = 0.6, 0.4 and 0.2 s by adding floor carpeting and absorptive panels on the walls and ceiling. More details on simulating reverberant conditions can be found in [10, 13].

2.2. Signal processing

2.2.1. Vocoder processing

This study implemented an eight-channel sinewave vocoder to simulate CI speech processing [11]. Briefly, MHINT sentences were processed through a pre-emphasis filter (with a 3 dB/octave roll-off and 2000 Hz cut off frequency), and then band-passed filtered into 8 bands between 80 and 6000 Hz using sixth-order Butterworth filters. The cutoff frequencies were [80, 221, 426, 724, 1158, 1790, 2710, 4050, and 6000 Hz] for the 8 band-pass filters. The temporal envelope of each spectral channel was extracted by full-wave rectification, followed by a low-pass 2-order Butterworth filter with a 400 Hz cutoff frequency. The envelope of each band was then compressed by the static or adaptive compression strategies in this study (see Fig. 1). The SEC used a fixed compression ratio for the whole amplitude envelope to confine its dynamic range to a pre-set value (e.g., 5 dB); while the AEC continuously varied its compression ratio frame by frame, but limiting the maximum and minimum values (or dynamic range) of the compressed amplitude within the pre-set range. The compressed envelopes then modulated a set of sinewaves with frequencies equal to the center frequencies of the band-pass filters. Finally, the envelope-modulated sinewaves of the 8 bands were summed up, and the level of the summed signal was adjusted to yield the same root-mean-square value of the original input signal.

2.2.2. Static envelope compression (SEC)

This study adopted a simple method to adjust the dynamic range of amplitude envelope, as proposed in [5]. Let x and y denote the input and output amplitude envelopes, respectively. The output compressed amplitude envelope y is computed, as:

$$y = \alpha \times (x - \bar{x}) + \bar{x}, \tag{1}$$

where $\bar{x}$ is the mean of the input amplitude envelope x, and α is a constant for compression factor, which is chosen in order for the output amplitude envelope to fall into a certain dynamic range DR, as:

$$UB = LB \times 10^{\frac{DR}{20}}, \tag{2}$$

where UB and LB stand for the upper bound and lower bound of output amplitude values, respectively. It is clear that the mean value of the output amplitude envelope equals the mean value of the input amplitude envelope (i.e., $\bar{y} = \bar{x}$), regardless the value of DR selected. It was found that by empirically setting compression factor α to 1/13, the dynamic range of multi-channel amplitude envelopes of MHINT sentences was adjusted to 5 dB [5].

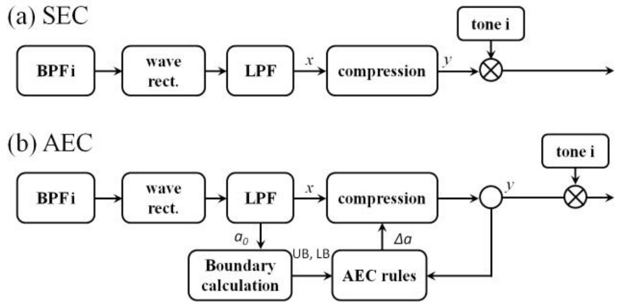

Fig. 1. Block diagrams to implement the (a) SEC- and (b) AEC-based speech processor in one channel [7].

Note that a small compression factor α denotes a large compression ratio, and vice verse. When α equals 0 in Eq. (1), the compressed amplitude envelope is a DC signal with a constant value of $\bar{x}$ (i.e., $\bar{y} = \bar{x}$), and the dynamic range is 0 dB; on the other hand, when α equals 1 in Eq. (1), the output amplitude envelope keep the original dynamic range of the input amplitude envelope. Figure 1 (a) shows the block diagram of the SEC-based speech processor. Note that a fixed compression factor a is applied here to the whole amplitude envelope to confine its dynamic range to a pre-set value.

2.2.3. Adaptive envelope compression (AEC)

Figure 1 (b) shows the implementation of the AEC-based speech processing. The AEC strategy includes two more units, i.e., boundary calculation and AEC rules. The compression factor α is initialized by a pre-defined value α_0 (the same as that used in SEC) to define a target dynamic range. The two boundaries (i.e., UB and LB) are first determined with the amplitude envelope and initial compression factor α_0, as:

$$\begin{cases} UB = \bar{x} + \alpha_0 \times (\max(x) - \bar{x}) \\ LB = \bar{x} + \alpha_0 \times (\min(x) - \bar{x}) \end{cases}, \tag{3}$$

where max (x) and min (x) denote the maximal and minimal values of input amplitude envelope x. Note that the UB and LB calculated in Eq. (3) are also present in the SEC-processed amplitude envelope. Nevertheless, instead of using the fixed compression factor α for all frames of the amplitude envelope, the AEC strategy here uses these two bounds to adjust the compression ratio for each frame (i.e., 2.5 ms in this study), as controlled by the AEC rules in Fig. 1 (b). Then the compressed amplitude envelope is computed as:

$$y_t = \alpha_t \times (x_t - \bar{x}) + \bar{x}, \tag{4}$$

where α_t is the adaptive compression factor that may increase or decrease the next one, as:

$$\alpha_{t+1} = \alpha_t + \Delta\alpha, \tag{5}$$

where $\Delta\alpha$ is determined by two AEC rules, as follows.

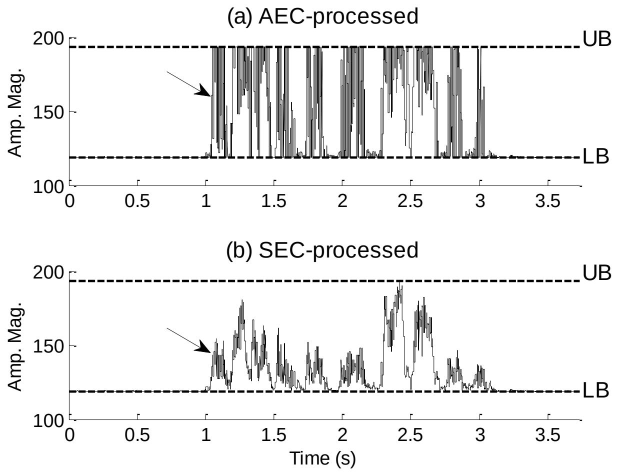

Fig. 2. Example of amplitude envelope processed by (a) AEC and (b) SEC strategies. The envelope is extracted from one channel of a testing sentence in reverberant condition T_{60}=200 ms, and compressed to 5 dB dynamic range within [LB, UB].

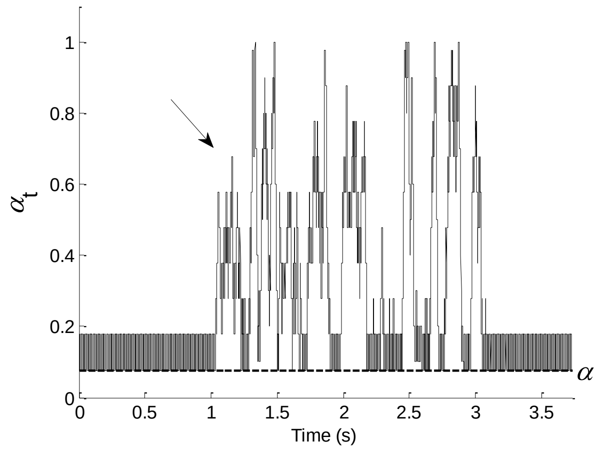

Fig. 3. The compression ratio α_t used in the AEC strategy for the compressed amplitude envelope in Fig. 2 (a). The dashed line denotes the level of fixed compression ratio α=1/13 used in SEC.

The first rule, i.e., the increasing envelope rule is designed to keep the compression processing as close to linear (i.e., α=1) as possible. By doing so, more signal can be pertained with less compression to produce the output signal. When y_t lies between UB and LB, AEC will increase α_t once by using a positive $\Delta\alpha$ (i.e., 0.1 in this study) in Eq. (5), and thus the compression ratio becomes larger. This increasing envelope rule stops if α_{t+1} reaches 1, whereas the original signal is used as the output signal without any compression. The second rule, i.e., the decreasing envelope rule is designed to ensure that the output envelope amplitude will not be out of the pre-set dynamic range [LB, UB]. When y_t is higher than UB or lower than LB, AEC will decrease α_t once by using a negative $\Delta\alpha$ (–0.1 in this study) in Eq. (5). This decreasing envelope rule stops if α_{t+1} reaches the initial value, i.e., α_0. More details on implementing the AEC strategy can be found in [7].

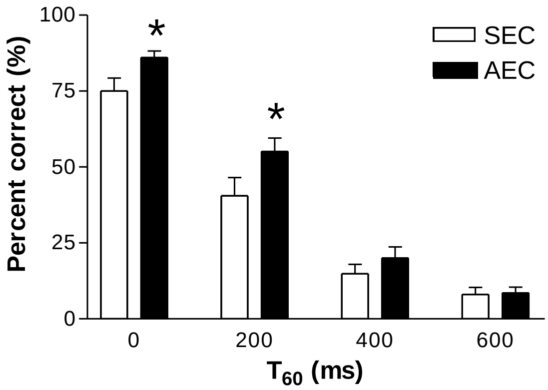

Fig. 4. Mean recognition scores of all conditions. The dynamic range of envelope amplitude is confined to 5 dB. The error bars denote the standard errors of mean values. Asterisk indicates a statistically significant (p<.05) difference between SEC and AEC scores.

Figure 2 (a) shows an example for adaptive envelope compression in one channel of vocoder processing. The initial compression factor α_0 is set to 1/13 to compress the envelope dynamic range to 5 dB (bounded by LB=118.7 and UB=193.5 in this example). SEC applies a fixed compression factor α =1/13 and AEC instead continuously adjusts its compression factor α_t, as shown in Fig. 3. Although different compression factors are applied to frames, the dynamic range is still confined to 5 dB in Fig. 2 (a), because the amplified amplitude envelope is still within the range limited by [LB, UB].

2.3. Procedure

All experiments in this study were conducted in a soundproof booth at the University of Hong Kong. The stimuli were played to listeners through a set of Sennheiser HD headphone at a comfortable listening level. The present work compressed the envelope dynamic range to 5 dB in vocoder simulation. This was done by using the compression factor α=1/13 and α_0=1/13 in Eqs. (1) and (4), respectively. Each subject participated in a total of 8 [= 4 reverberant conditions (i.e., T_{60}=0, 200, 400, and 600 ms) × 2 envelope compression strategies (i.e., SEC and AEC)] testing conditions. Each condition contained 10 sentences. The order of the 8 conditions was randomized across subjects, and none of the 10 sentences were repeated across testing conditions. Subjects repeated what they heard during the experiment, and were allowed to repeat the stimuli twice. Sentence recognition score was calculated by dividing the number of words correctly identified by the total number of words (i.e., 100) in each testing condition.

3. RESULTS

Mean sentence recognition scores for all conditions are shown in Fig. 4. Statistical significance was determined by using the percent recognition score as the dependent variable, and reverberant condition (i.e., T_{60} value) and compression strategy as the two within-subject factors. Two-way analysis of variance with repeated measures indicated a significant effect (F[3, 24]=208.62, p<.005) of reverberant condition, compression strategy (F[1, 8]=35.01, p<.005), and a significant interaction (F[3, 24]=8.05, p=.001) between reverberant condition and compression strategy. Post hoc analysis showed that for paired score comparison (i.e., SEC-processed vs. AEC-processed), the score differences between SEC-processed sentences and AEC-processed sentences were significantly (p<.05) at reverberant conditions of T_{60}=0 and 200 ms in Fig. 4.

4. DISCUSSION AND CONCLUSIONS

Following the previous study assessing the contribution of adaptive envelope compression for speech understanding in noise, the present work further examined its effect for reverberated speech. The AEC strategy adaptively modified compression ratio based on the characteristic of the input speech signal, and optimally utilized the usable dynamic range. Using a small compression ratio could provide larger dynamic range of amplitude envelope, thus yielding a large modulation depth of amplitude envelope. Consistently with the intelligibility advantage observed from the previous study [7], the present work showed that the amplitude envelope processed by the AEC strategy led to a higher intelligibility for vocoded sentences in reverberation than that processed by the SEC strategy. This makes the AEC strategy a highly promising way to enhance speech understandings in reverberation for implanted listeners in the future.

It is also noted that the above intelligibility advantage was not observed at all reverberant conditions in the present work. That is, no significant intelligibility improvement was found at reverberant conditions of T_{60}=400 and 600 ms in Fig. 4. This may be partially attributed to the usage of initial compression parameters (e.g., α_0 and $\Delta\alpha$) in this study. Further study is warranted to study the effect of optimal compression parameters to the performance of the AEC-based speech processing in reverberation.

5. ACKNOWLEDGEMENTS

This work was partially supported by the National Science Council of Taiwan under contract NSC101-2221-E-001-020- MY3. This research was also supported by Faculty Research Fund (Faculty of Education) and Seed Funding for Basic Research, The University of Hong Kong.

6. REFERENCES

[1] P. C. Loizou, “Introduction to cochlear implants,” IEEE Eng. Med. Biol. Mag., vol. 18(1), pp. 32–42, 1999.

[2] F.G. Zeng, “Trends in cochlear implants,” Trends Amplif., vol. 8(1), pp. 1–34, 2004.

[3] F.G. Zeng, G. Grant, J. Niparko, J. Galvin, R. Shannon, J. Opie, and P. Segel, “Speech dynamic range and its effect on cochlear implant performance,” J. Acoust. Soc. Am., vol. 111(1), pp. 377–386, 2002.

[4] R. van Hoesel, M. Böhm, R.D. Battmer, J. Beckschebe, and T. Lenarz, “Amplitude-mapping effects on speech intelligibility with unilateral and bilateral cochlear implants,” Ear Hear., vol. 26(4), pp. 381–388, 2005.

[5] F. Chen, L.L. Wong, J. Qiu, Y. Liu, B. Azimi, and Y. Hu, “The contribution of matched envelope dynamic range to the binaural benefits in simulated bilateral electric hearing,” J. Speech, Lang. Hear. Res., vol. 56(4), pp. 1166–1174, 2013.

[6] P.J. Blamey, “Adaptive dynamic range optimization (ADRO): A digital amplification strategy for hearing aids and cochlear implants,” Trends Amplif., vol. 9(2), pp. 77–98, 2005.

[7] C.J. James, P.J. Blamey, L. Martin, B. Swanson, Y. Just, and D. Macfarlane, “Adaptive dynamic range optimization for cochlear implants: A preliminary study,” Ear Hear., vol. 23(1), pp. 49S–58S, 2002.

[8] Y.H. Lai, F. Chen, and Y. Tsao, “An adaptive envelope compression strategy for speech processing in cochlear implants,” in Proc. of 15th Annual Conference of the International Speech Communication Association, Singapore, 2014, in press.

[9] Y.H. Lai, P.C. Li, K.S. Tsai, W.C. Chu, and S.T. Young, “Measuring the long-term SNRs of static and adaptive compression amplification techniques for speech in noise,” J. Am. Acad. Audiol., vol. 24(8), pp. 671–683, 2013.

[10] F. Chen, O. Hazrati, and P.C. Loizou, “Predicting the intelligibility of reverberant speech for cochlear implant listeners with a non-intrusive intelligibility measure,” Biomedical Signal Processing & Control, vol. 8 (3), pp. 311–314, May 2013.

[11] F. Chen, and P.C. Loizou, “Predicting the intelligibility of vocoded and wideband Mandarin Chinese,” J. Acoust. Soc. Am., vol. 129(5), pp. 3281–3290, 2011.

[12] L.L. Wong, S. Soli, S. Liu, N. Han, and M. Huang, “Development of the Mandarin hearing in noise test (MHINT),” Ear Hear., vol. 28, pp. 70S–74S, 2007.

[13] T. Van den Bogaert, S. Doclo, J. Wouters, and M. Moonen, “Speech enhancement with multichannel Wiener filter techniques in multi-microphone binaural hearing aids,” J. Acoust. Soc. Am., vol. 125 (1), pp. 360–371, 2009.