An Adaptive Envelope Compression Strategy for Speech Processing in Cochlear Implants

Ying-Hui Lai [1], *Fei Chen*[2], *Yu Tsao*[1]

[1] Research Center for Information Technology Innovation, Academia Sinica, Taipei, Taiwan
[2] Division of Speech and Hearing Sciences, The University of Hong Kong, Hong Kong, China

jackylai@citi.sinica.edu.tw, feichen1@hku.hk, yu.tsao@citi.sinica.edu.tw

Abstract

Hearing-impaired patients have limited hearing dynamic range for speech perception, which partially accounts for their poor speech understanding abilities, particularly in noise. Wide dynamic range compression aims to compress speech signal into the usable hearing dynamic range of hearing-impaired listeners; however, it normally uses a static compression based strategy. This work proposed a strategy to continuously adjust the envelope compression ratio for speech processing in cochlear implants. This adaptive envelope compression (AEC) strategy aims to keep the compression processing as close to linear as possible, while still confine the compressed amplitude envelope within the pre-set dynamic range. Vocoder simulation experiments showed that, when narrowed down to a small dynamic range, the intelligibility of AEC-processed sentences was significantly better than those processed by static envelope compression. This makes the proposed AEC strategy a promising way to improve speech recognition performance for implanted patients in the future.

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

1. Introduction

Cochlear implant (CI) is a surgically implanted electronic device that provides a sense of sound to a person with profound-to-severe hearing loss [1-2]. In a CI device, the input signal is received by a microphone and fed into a speech processor. The speech processor captures the multi-channel temporal envelopes of input 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 real speech signal. Hence, a compression scheme is required to compress the DR of 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]. For instance, Zeng et al. and van Hoesel et al. found that with a lower DR, the speech recognition performance was degraded in both quiet and noisy conditions [3-4].

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, adaptive dynamic range control strategies were developed. For instance, the adaptive dynamic range optimization (ADRO) method selected the most information-bearing segments from input signal, and presented them to hearing-impaired listeners for recognition. ADRO has shown notable intelligibility and quality improvement reported from hearing aids (HA) and CI users [6-7]. Wide-dynamic-range compression (WDRC) amplification scheme is another well-known strategy being widely used in present HA and CI devices. It uses static compression ratio to satisfy loudness requirement over a wide input levels [8]. Although WDRC provides satisfactory performance in quiet, studies reported that it could produce unsatisfactory speech intelligibility and sound quality in noisy conditions [e.g., 9]. Recently, Lai et al. proposed an adaptive WDRC (i.e., AWDRC) strategy, which dynamically adjusted compression ratio according to the short-term dynamic range of input signal. Experimental results indicated that AWDRC could provide better long-term SNR scores than the conventional method that used a static compression ratio [10-11].

Following the development of the AWDRC strategy, the purpose of this study is to propose an adaptive envelope compression (AEC) strategy to improve speech understanding in noise for cochlear implants. More specifically, this study will compare the recognition of speech synthesized by the proposed AEC strategy and the static envelope compression (SEC) strategy. To assess the effect of dynamic range compression to the speech perception in noise, 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 [12].

2. Methods

2.1. Subjects and materials

Eleven (18-24 yrs., 6 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 [13]. All sentences were pronounced by a male native-Mandarin speaker, with fundamental frequency ranging from 75 to 180 Hz, and recorded at a sampling rate of 16 kHz. Two types of maskers, i.e., speech-shaped noise (SSN) and two equal-level interfering male talkers (2T) were used to corrupt test sentences at two signal-to-noise ratio (SNR) levels of 10 and 5 dB, which were chosen to avoid the ceiling/floor effects.

2.2. Signal processing

2.2.1. Vocoder processing

This study adopted an 8-channel sinewave vocoder to simulate cochlear implant speech processing [12]. MHINT sentences were first processed through a pre-emphasis filter (with a 3 dB/octave roll-off and 2000 Hz cut off frequency), and then band-passed filtered (BPF) into eight frequency 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 channel allocation of band-pass filters. The temporal envelope of each spectral channel was extracted by full-wave rectification followed by a low-pass second-order Butterworth filter (LPF) with a cutoff frequency of 400 Hz. The envelope of each band was then compressed by static and adaptive compression strategies in this study (see Fig. 1, and more on this later). The static envelope compression used a fixed compression ratio for the whole amplitude envelope to confine its dynamic range to a pre-set value; while the adaptive envelope compression 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 eight 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

Several methods have been proposed to adjust the dynamic range of amplitude envelope. This study adopted a simple method 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 (i.e., compression factor) 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 are the upper bound (i.e., maximum) and lower bound (i.e., minimum) 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. Chen et al. found that by empirically setting compression factor α to 1/3, 1/5, and 1/13, the dynamic ranges of multi-channel amplitude envelopes of MHINT sentences were adjusted to 15, 10, and 5 dB, respectively [5].

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 becomes 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 in one channel. 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

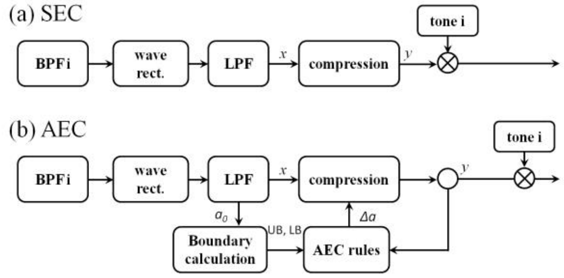

Figure 1. *Block diagrams of implementing the (a) SEC- and (b) AEC-based speech processor in one channel.*

Figure 1 (b) shows the block diagram of the AEC-based speech processor in one channel. Comparing Figs. 1 (a) and (b) reveals that 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) are 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 (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:

[1] <u>Increasing envelope rule:</u> This rule is 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.

[2] <u>Decreasing envelope rule:</u> This rule ensures 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 $\Delta\alpha$ (i.e., –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.

In addition, the amplitude of the compressed envelope [controlled by α_t in Eq. (4)] is set to UB or LB if it is larger than UB or smaller than LB. Figure 2 (a) shows an example for adaptive compression for amplitude envelope in one channel. The initial compression factor α_0 is set to 1/13 (i.e., to compress the envelope dynamic range to 5 dB). The UB and LB in this example are 193.3 and 277.6, respectively, yielding a dynamic range around 5 dB. Both SEC and AEC aim to

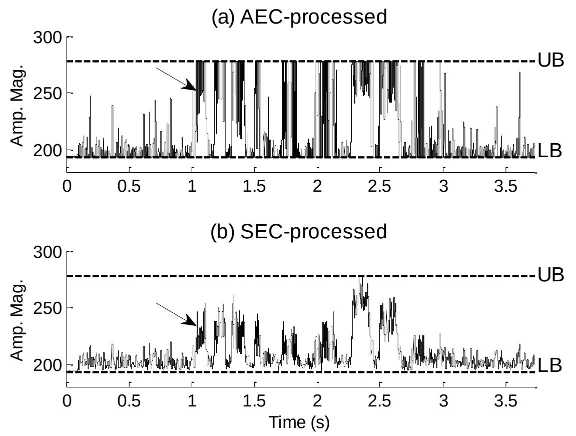

Figure 2. *Example of amplitude envelope processed by (a) AEC and (b) SEC strategies. The envelope is extracted from the 6th channel of a testing sentence masked by SSN at 5 dB SNR, and compressed to 5 dB dynamic range within [LB, UB].*

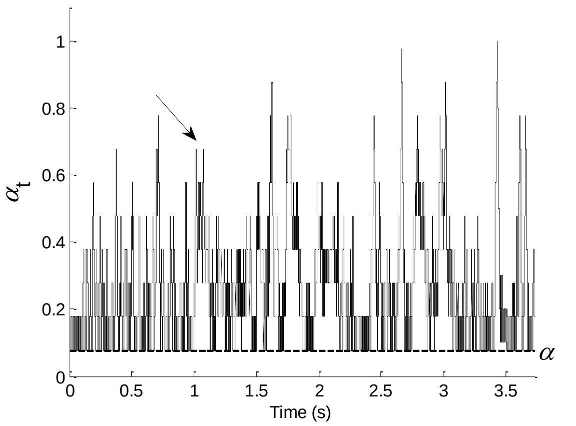

Figure 3. *The compression ratio α_t used in the AEC strategy for the compressed amplitude envelope in Fig. 2 (a). The dashed line denotes the level of fixed compression ratio α=1/13 in SEC.*

compress the amplitude envelope within this range, as shown in Figs. 2 (a) and (b); however, 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). This is because the amplified amplitude envelope is still within the range limited by [LB, UB].

2.3. Procedure

The experiment was conducted in a soundproof booth at the University of Hong Kong. The stimuli were played to listeners through a set of Sennheiser HD headphones at a comfortable listening level. This study 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 8 [= 2 SNR levels × 2 types of maskers × 2 envelope compression strategies (SEC and AEC)] testing conditions. Each condition contained 10 sentences, and the order of the 8 conditions was randomized across subjects. None of the 10 sentences were repeated across testing conditions. Subjects were instructed to repeat what they heard, 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 in each testing condition. During testing, each subject was given a 5-min break every 30-min during the test.

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 SNR level and compression strategy as the two within-subject factors. The scores were first converted to rational arcsine units (RAU) using the rationalized arcsine transform.

For conditions tested with SSN masker in Fig. 4 (a), two-way analysis of variance (ANOVA) with repeated measures indicated a significant effect (F[1, 10]=36.03, p<.005) of SNR level, compression strategy (F[1, 10]=33.41, p<.005), and a significant interaction (F[1, 10]=5.52, p=.041) between SNR level and compression strategy. Post hoc analysis showed that for both 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) in Fig. 4 (a).

For conditions tested with 2T masker in Fig. 4 (b), two-way ANOVA with repeated measures indicated a significant effect (F[1, 10]=77.87, p<.005) of SNR level, a non-significant effect of compression strategy (F[1, 10]=3.14, p=.107), and a significant interaction (F[1, 10]=12.63, p=.005) between SNR level and compression strategy. Post hoc analysis showed that for paired score comparison (i.e., SEC-processed vs. AEC-processed), the score difference at 5 dB SNR was significantly (p<.05) while that at 10 dB SNR was non-significant (p=0.46) in Fig. 4 (b).

4. Discussion and conclusions

This study proposed an adaptive envelope compression strategy for CI speech processing. The AEC strategy adaptively modified compression ratio based on the characteristic of input signal, and optimally utilized the usable dynamic range. Using a small compression ratio [or a large compression factor α_t in Eq. (4)] could provide larger dynamic range of amplitude envelope, thus yielding a large modulation depth of amplitude envelope. This is manifested by the compressed amplitude envelope exemplified in Fig. 2, whereas the AEC-processed envelope around 1 sec. in Fig. 2 (a) has a larger dynamic range than that processed by the SEC strategy in Fig. 2 (b). Figure 3 also shows that a small amplitude compression ratio (or a large compression factor α_t) is used for this frame, in contrast to the fixed compression factor (i.e., α=1/13) employed in the SEC strategy. Previous studies suggested that modulation depth was an important factor to speech perception, especially under noisy conditions [14]. The improved modulation depth may partially account for the better intelligibility of vocoded sentences synthesized with AEC-processed envelopes.

The signal processing in the AEC strategy is similar to that in the SEC process, but is characterized with additional boundary calculation (for UB and LB) and AEC rules to optimally and continuously adjust compression ratio on a frame basis. Since these two additional units are rather simple,

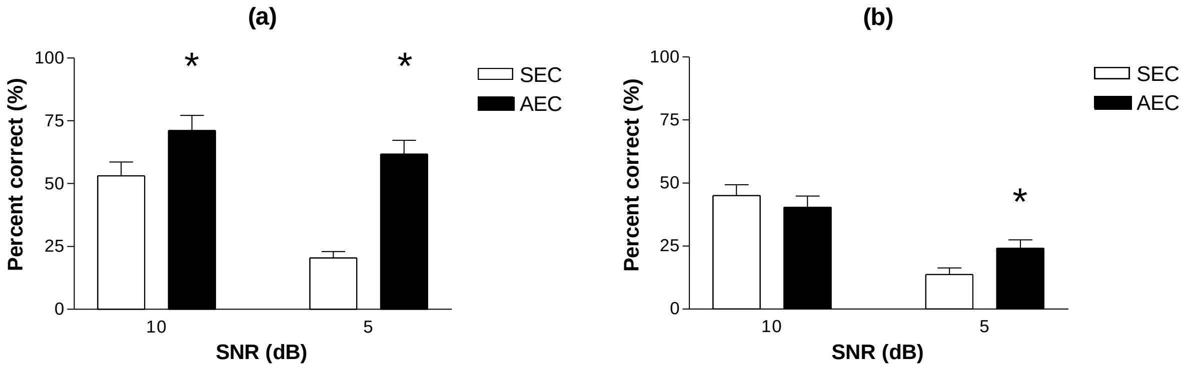

Figure 4. *Mean recognition scores of all conditions at (a) SSN and (b) 2T maskers. The dynamic range of envelope amplitude is confined to 5 dB. The error bars denote the standard errors of mean values. Asterisk indicates a statistically significant (p<.05) difference between SEC and AEC scores.*

the computation load for AEC is reasonable compared to the conventional SEC processing, making AEC feasible to be implemented using microprocessors.

A non-linear compressive mapping function is normally used in CI devices to convert acoustic amplitude envelope to electric current signal (with a narrow dynamic range). As the present study assessed the performance of compression strategy (i.e., static vs. adaptive) to CI speech processing by vocoder simulation, we used a simple compression function [i.e., Eqs. (1) and (4)] to compress amplitude envelope into a pre-set dynamic range. Note that most of the present acoustic-to-electric conversions in CI devices use fixed mapping function. It is reasonable to foresee that the adaptive mapping (from acoustic to electric) function will improve the speech understanding of implanted patients. Further studies will be conducted to incorporate the AEC strategy proposed in this study with the present CI processors, and evaluate its performance with CI patients.

In conclusion, the present work proposed an adaptive envelope compression strategy to confine the amplitude envelope within a fixed dynamic range, but continuously adjust compression ratio for short-term amplitude. Through this processing, the local dynamic range approached to that of the uncompressed amplitude envelope. Consistently with the intelligibility advantage observed from previous studies, the present work showed that the amplitude envelope processed by the AEC strategy led to a significantly high intelligibility for vocoded sentences in noise than that processed by the SEC strategy. This makes the proposed AEC strategy a highly promising way to enhance speech understandings in noise for implanted listeners in the future.

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] Loizou, P.C., "Introduction to cochlear implants," IEEE Eng. Med. Biol. Mag., 18(1): 32-42, 1999.

[2] Zeng, F.G., "Trends in cochlear implants," Trends Amplif., 8(1): 1-34, 2004.

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

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

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

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

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

[8] Rosengard, P.S., Payton, K.L. and Braida, L.D., "Effect of slow-acting wide dynamic range compression on measures of intelligibility and ratings of speech quality in simulated-loss listeners," J. Speech, Lang. Hear. Res., 48(3): 702-714, 2005.

[9] Naylor, G. and Johannesson, R.B., "Long-term signal-to-noise ratio at the input and output of amplitude-compression systems," J. Am. Acad. Audiol., 20(3): 161-171, 2009.

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

[11] Lai, Y.H., Tsao, Y. and Chen, F. "A study of adaptive WDRC in hearing aids under noisy conditions," Int. J. Speech Lang. Pathol. Audiol., 1(2), 43-51, 2013.

[12] Chen, F. and Loizou, P.C., "Predicting the intelligibility of vocoded and wideband Mandarin Chinese," J. Acoust. Soc. Am., 129(5): 3281-3290, 2011.

[13] Wong, L.N., Soli, S., Liu, S., Han, N. and Huang, M., "Development of the Mandarin hearing in noise test (MHINT)," Ear Hear., 28:70S-74S, 2007.

[14] Souza, P.E. and Boike, K.T., "Combining temporal-envelope cues across channels: Effects of age and hearing loss," J. Speech, Lang. Hear. Res., 49(1): 138-149, 2006.