Speech Intelligibility Evaluation of Sound Coding Strategies in Noisy Environments for Thai Cochlear Implant Users

Siriporn Dachasilaruk[1]* and Apichai Rungruang[2]

[1]Department of Electrical and Computer Engineering, Faculty of Engineering, Naresuan University, Phitsanulok, 65000, Thailand

[2]Department of English Language, Faculty of Humanities, Naresuan University, Phitsanulok, 65000, Thailand

* Corresponding author. E-mail address: sirippornd@nu.ac.th, dsiriporn@hotmail.com

Abstract

The paper presents the effect of sound coding strategies on speech intelligibility of Thai-speaking cochlear implant (CI) users. Two sound coding strategies, namely continuous interleaved sampling (CIS) and advanced combination encoder (ACE) strategies were evaluated for word recognition. The tested words consisting of monosyllabic and bisyllabic words were corrupted by speech shaped noise and babble noise at SNR levels of 0, 5, 10 and 15 dB. The vocoded speech of clean and noisy words were tested by twelve normal-hearing listeners. The experimental results showed that speech intelligibility of the ACE strategy had higher mean scores than that of the CIS strategy in all tested conditions. The ACE strategy provided a significant speech intelligibility at SNR levels of 5 and 10 dB for monosyllabic words with speech shaped noise and at SNR levels of 0 dB for bisyllabic words with speech shaped noise. In addition, the noisy bisyllabic words provided higher intelligibility performance than the noisy monosyllabic words in all tested conditions.

Keywords: cochlear implant, sound coding, speech intelligibility, filter bank, hearing loss

Introduction

An important problem of hearing-impaired (HI) people is communication with other persons in daily life, in particular in noisy environments. Although hearing devices (e.g. hearing aids (HAs) and cochlear implants (CIs)) can help them to improve their auditory and communicative ability, there are still many limitations which relate to user-related factors and device-related factors. For example of user-related factors, either HA users or CI users may lose hearing at the same degrees but they may have different abilities to understand speech. In addition, CI users relate to pre- and post-lingual deafened CI users, deafness duration and so on. For example, device-related factors, HAs are amplified acoustic devices whereas CIs are electric acoustic devices. Either HA users or CI users with the same degrees of hearing loss may use the same devices but their speech intelligibility may be different. Typically, CIs are an effective treatment for HI people with severe to profound hearing loss who cannot benefit from HAs. However, CIs are limited by frequency, temporal and amplitude resolutions (Qin & Oxenham, 2003) which affect transmitting the speech information to the auditory nervous system.

CI systems have been developed to mimic function of the normal human cochlea over thirty decades (Fan-Gang, Rebscher, Harrison, Xiaoan, & Haihong, 2008; Loizou, 2006; Wilson & Dorman, 2008). CIs have effectively restored hearing to more than 320,000 people worldwide as of December 2012 and unpublished industry records have revealed that the number of CI users exceed half a million in early 2016 (Wilson, 2017). The most important part of the CI system is the sound coding strategy which converts the speech sound to electrical stimulation and maximizes the user's communicative capability. Several sound coding strategies were proposed by CI manufacturers. These strategies can divided into two main types (Loizou, 2006). One is an implementation of all envelope channels selected for stimulation, such as Continuous Interleaved Sampling (CIS) and HiResolution (HiRes) strategies. The other is an implementation of only a subset of the envelope channels (*n*-of-*m*) selected for stimulation, such as Spectral Peak (SPEAK) and Advanced Combination Encoder (ACE) strategies. The CIS strategy is commonly specified as an option in all CI systems (Fan-Gang et al., 2008; Wilson & Dorman, 2008) and the other strategies are specified as default strategies in each CI system of manufacturers.

Generally, the main stages of these strategies include a bank of bandpass filter, envelope detection, compression and modulation. These stages are designed by different methods depending on cochlear devices of each manufacturer (Loizou, 2006). For example, a filter bank of some devices is implemented by finite impulse response (FIR) filters, infinite impulse response (IIR) filters or fast Fourier transform (FFT). The envelope detection in each channel of different devices are estimated by using full-wave rectification and lowpass filtering, or the Hilbert transform. In addition, parameters (e.g. filter spacing and cutoff frequencies) in each stage are varied to optimize speech intelligibility.

The performance of speech intelligibility for these strategies has been widely reported. Nevertheless, the majority of coding strategies have been designed and evaluated using the English language that differs from Thai language. Different languages based on the same coding strategy may yield different performance of speech intelligibility due to different acoustic cues and phonemics. English is a stress-timed language. Stress patterns play a crucial role in meanings and grammatical functions. For example, the term “record” is a noun when the stress in on the first syllable; it can be a verb when the stress is placed on the second syllable. Needless to say, the meanings are different as well. That is, one is a tool when “record” functions as a noun. Another means to store sounds when it is a verb. On the other hand, tonal languages like Thai or Chinese use tone (pitch) variations to identify the meaning differences.

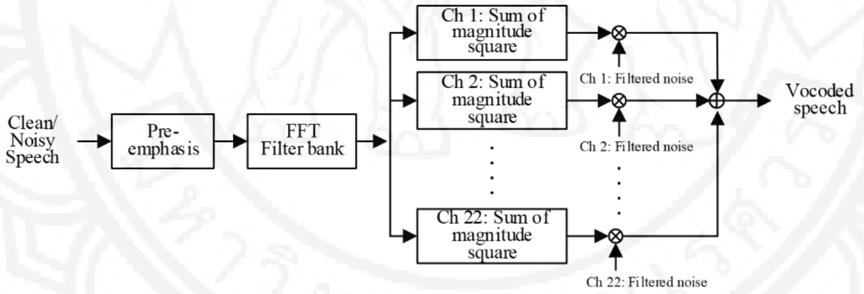
Normally, Thai syllables have five lexical tones, namely the middle / ¯/, the low /`/, the falling /ˆ/, the high /´/, and the rising /ˇ/. For instance, in Thai the term “ma” in a middle tone means to come. It becomes a dog or a horse when it is in a rising tone and a high tone, respectively. In Chinese, Lin (Lin, 2013) stated that the term “ma” in a high falling tone means to scold, but it refers to a horse in a low-falling-rising tone. In other words, the tone changes the meanings. Although Thai and Chinese (e.g. Mandarin) are tonal languages, they have different acoustic cues. For example, Thai has five tones with a range of approximately 50-200 Hz while Mandarin Chinese has four tones with a range of approximately 50-350 Hz (Xu & Zhou, 2011). These components are important to specifically investigate the performance of sound coding strategies among Thai CI users.

In Thailand, the first case of CI was performed in 1986 (Vaewvichit & Luangpitakchumpol, 1999). Several sound coding strategies were used by Thai CI users (Kasemsuwan et al., 2010; Vaewvichit & Luangpitakchumpol, 1999). However, these strategies have never been evaluated and compared using the Thai language in noisy listening environments. As mentioned above, the CIS and ACE were evaluated and compared in the current study. Both strategies are commercially manufactured by Cochlear Corporation and the ACE is widely used by the majority of CI users (Wouters, McDermott, & Francart, 2015). This paper is organized in the following way. The strategies of the CI sound coding are presented in section 2. The performance evaluation is described in terms of subjects and procedures in section 3. The results and discussion are explained in section 4. Finally, the conclusion is given in section 5.

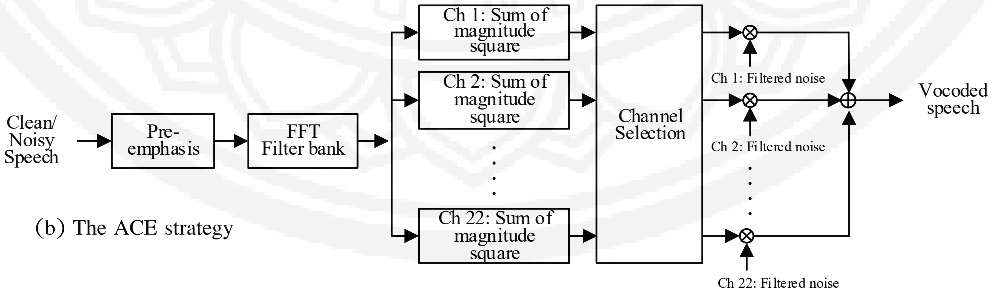
Cochlear Implant Sound Coding Strategies

The CIS and the ACE strategies in Nucleus CI devices transmit only temporal-envelope information as patterns of the electrical stimulation which contribute to speech intelligibility (Fan-Gang et al., 2008). Both strategies are implemented with the FFT for a bank of bandpass filters. However, the stage of the channel selection is only implemented in the ACE but not in the CIS as in Figure 1. A subset of the envelope channels with the largest amplitudes is selected for simultaneous stimulation. This concept is also called the *n*-of-*m* strategy. The advantages of the *n*-of-*m* strategy is to increase temporal resolution (stimulation rate) which can update frequently the important information of speech (Nogueira, Buchner, Lenarz, & Edler, 2005). At the same time, this strategy can reduce redundant information of speech by deleting the envelope channels with the small amplitudes which are the less significant channels. This can probably reduce electrode interaction in CI devices (Wilson & Dorman, 2008).

Generally, the CI simulation has been widely used in the first step of evaluation to study performance of CI users and to avoid the confounding factors depending on CI users, such as duration of implantation, surviving neural pattern, or individual preference (Chen & Loizou, 2011; Loizou, 2006). The speech signal is simulated to generate the vocoded speech similar to CI processors in cochlear devices and present to normal-hearing (NH) listeners. The block diagram of the CI simulation for the ACE with a noise-band vocoder is presented in Figure 1(b). The clean or noisy speech are sampled with sampling rate of 16 kHz. The high-frequency components of the clean or noisy speech were amplified by a pre-emphasis filter. After that, the pre-emphasized speech is performed by frame-based processing with a window of 128 samples and an overlap of 75%.



(a) The CIS strategy



(b) The ACE strategy

Figure 1 Block diagram of the CI simulation based on a noise-band vocoder

In each frame, the pre-emphasized speech is transformed with the FFT into frequency subbands (128 bins with 125 Hz width). Only the first 64 bins are combined by summing the power of consecutive bins to produce 22 channels (electrodes). The cutoff frequencies and the center frequencies of 22 channels (Cochlear, 2002) are allocated as follows. The cutoff frequencies are 187.5, 312.5, 437.5, 562.5, 687.5, 812.5, 937.5, 1062.5, 1187.5, 1312.5, 1562.5, 1812.5, 2062.5, 2312.5, 2687.5, 3062.5, 3562.5, 4062.5, 4687.5, 5312.5, 6062.5, 6937.5, and 7937.5 Hz. The center frequencies are 250, 375, 500, 625, 750, 875, 1000, 1125, 1250, 1437, 1687, 1937, 2187, 2500, 2875, 3312, 3812, 4375, 5000, 5687, 6500, and 7437 Hz. The frequency spacing is linearly spaced below 1 kHz and logarithmically spaced above 1 kHz. The power of the envelope is weighted by a weighting value of each channel. Then, a 12-channel envelope with the largest amplitude is selected and modulated by white noise with the same cutoff frequencies of 22 channels. All the selected channels of the modulated speech are summed to synthesize the vocoded speech. The vocoded speech as a simulated sound of the CI system is presented to the NH listeners.

Performance Evaluation

Subjects

Twelve normal-hearing subjects (ten males and two females, age range from 22 to 34 and mean age 24) were recruited in this experiment. All subjects were native Thai speakers, and they were undergraduate students and staff at a public university. They had normal hearing thresholds (<25 dB HL, between 0.25 and 8 kHz). The experiment was approved by the Ethics Committee of the university.

Procedures

All the Thai words were selected from a corpus used in clinical practice with HI listeners. All words were recorded by a male Thai speaker at a sampling rate of 22 kHz. The Thai word test consisted of 8 lists of Thai monosyllabic words and 8 lists of Thai bisyllabic words (Kangsanarak, 1980; Komalarajun, 1979). Each list had 25 words (16 lists × 25 words = 400 words). An example of monosyllabic and bisyllabic words are shown in Table 1. These words were corrupted by speech-shaped noise (SSN) at 0, 5 and 10 dB SNR, and babble noise (BBN) at 5, 10 and 15 dB SNR. The SNR levels were carefully chosen to avoid floor and ceiling effects (Chen, Hu, & Yuan, 2015), especially the noisy monosyllabic words. The clean and noisy words were simulated using the CIS and the ACE strategies to produce the vocoded speech in a total of 28 conditions (2 coding strategies × 2 word types × 2 noise types × 3 SNR levels for the noisy words and 2 coding strategies × 2 word types for the clean words).

The experiment was carried out in a soundproof room. The vocoded speech was presented using a laptop via a headphone. The subjects used only one preferred ear to listen to the vocoded speech in all tested conditions. The volumes of the vocoded speech were adjusted to be at a comfortable conversional level. A total of 28 conditions were evaluated by each subject over two sessions (14 conditions per session, one session per day), with a break of at least one week between the two sessions to avoid learning effects. Experiment in each session lasted approximately one hour.

Table 1 An example of monosyllabic and bisyllabic words

Monosyllabic words			Bisyllabic words		
Thai script	Romanization	Translation	Thai script	Romanization	Translation
เรือ	rue	ship	รอรถ	ro-rót	wait for the bus
ม้า	má	horse	เท่ากัน	thaô-kan	equal
ญาติ	yât	relative	เล็กน้อย	lék-noǐ	little
เดาะ	dò	slight kick	ตลาด	ta-làt	market
ว่า	wâ	blame	งานบ้าน	ngan-bân	house chore
ฟ้า	fá	sky	เอาใจ	ao-chai	please (verb)
ฉาย	chaǐ	light (verb)	ของใคร	không-khai	whose stuff is this?
แก่	kàe	old	ทุกเช้า	thúk-chaǒ	every morning
ยาม	yam	guard	นอนแล้ว	non-laěw	went to bed
กับ	kàp	with	ด้วยกัน	dûay-kan	together

Each subject undertook training tests before the real tests in both sessions in order to ensure that the subjects clearly understood how to do the tests and familiarise themselves with the vocoded speech. In the training tests the subjects were asked to listen to the vocoded speech in all tested conditions during a 5-minute test. In the real tests, the subjects were asked to write down the words they heard in the prepared papers. If they were uncertain, they could guess the words. The scores of subjects were counted based on how many words they identified correctly in each condition. In each session, no word list was repeated across the conditions and the order of tested conditions was randomized for each subject. During the real test, the subjects were given a break every 20 minutes, or whenever they required a rest, to avoid listening fatigue which may affect their performance.

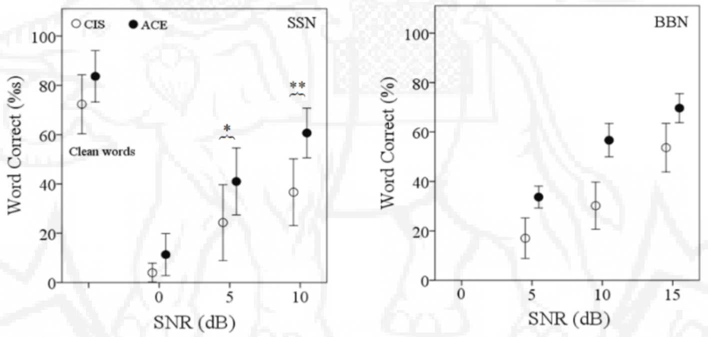
Results and Discussion

The mean percentage correct scores of the 12 NH subjects in 28 conditions are presented in Fig. 2. The tested conditions consisted of clean and noisy words of monosyllabic and bisyllabic words. The noisy words depended on noise types and SNR levels. The words of each condition were scored based on the percentage of words correct for each subject. Then, the scores of all subjects were averaged in each condition. The experimental results showed that mean scores of noisy words increased as SNR levels increased at the same noise type and word type. When compared with the monosyllabic words as in Fig. 2, the bisyllabic words yielded extremely high mean scores and their mean scores were almost the same as mean scores of clean words at high SNR levels. The mean scores of the ACE were remarkably higher than those of the CIS in all conditions.

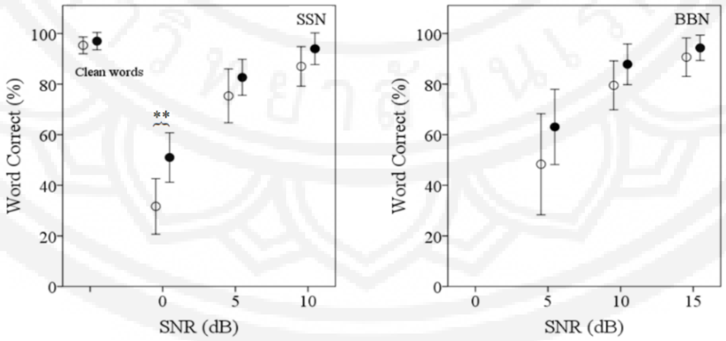
A two-way analysis of variance (ANOVA) with repeated measures was used to determine the relationship between the means of two factors, namely coding strategy and SNR level. A significant difference of mean scores was found in all SNR levels for both noise type and word type. This means the increased SNR levels yielded significantly mean intelligibility scores in all word types and noise types. For the monosyllabic words with SSN, the analysis revealed a significant effect of the coding strategy [$F(1,11)=81.93$, $p<0.0005$], a significant effect of the SNR level [$F(2,22)=116.78$, $p<0.0005$] and a significant interaction effect between the coding strategy and the SNR level [$F(2,22)=4.43$, $p<0.05$]. Post hoc tests of the coding strategy and the SNR level indicated that the ACE produced significantly higher intelligibility scores than the the CIS at the SNR level of 5 and 10 dB. For the monosyllabic words with BBN, there were significant effects of the coding strategy

[$F(1,11)=51.19$, $p<0.0005$] and the SNR level [$F(2,22)=103.08$, $p<0.0005$]. For the bisyllabic words with SSN, the analysis indicated a significant effect of the coding strategy [$F(1,11)=20.34$, $p<0.05$], a significant effect of the SNR level [$F(2,22)=262.53$, $p<0.0005$] and a significant interaction effect between the coding strategy and the SNR level [$F(2,22)=6.75$, $p<0.05$]. Post hoc tests of the coding strategy and the SNR level indicated that intelligibility perforamnce of the ACE was significantly higher than that of the CIS at 0 dB SNR. For the bisyllabic words with BBN, the analysis indicated a significant effect of the coding strategy [$F(1,11)=12.62$, $p<0.05$] and the SNR level [$F(2,22)=278.27$, $p<0.0005$].

The findings of the current study were consistent with those of previous studies. For example, the studies of Kiefer, Hohl, Sturzebecher, Pfennigdorff, and Gstoettner (2001) and Skinner et al. (2002) found that most Nucleus CI users preferred the ACE over the CIS. The intelligibility scores of the ACE were significantly higher than those of the CIS. Because of selecting a subset of the envelope channels with the largest amplitude, the ACE strategy can remove the less significant channels (e.g. noise) (Nogueira et al., 2005) and reduce the overall noise levels of noisy words. In case of the word type, the current study supports the experimental results of Vaewvichit and Luangpitakchumpol (1999). They found that nine Nucleus CI users with the bisyllabic words can achieve higher speech intelligibility than the monosyllabic words in quiet conditions. Moreover, most Nucleus CI users with lip reading provided higher speech intelligibility than that without lip reading.



(a) Monosyllabic words



(b) Bisyllabic words

Figure 2 Mean percentage correct scores of 12 NH listeners as a function of the CI coding strategies and the SNR levels for monosyllabic and bisyllabic words. The mean scores for the clean words and the noisy words are demonstrated for the speech-shaped noise (SSN) at different SNR levels (i.e. 0, 5, 10 dB) and the babble noise (BBN) at different SNR levels (i.e. 5, 10, 15 dB). An asterisk (*) denotes that the mean scores are significantly higher performance than those at the same SNR level; $^{*}p < 0.05$, $^{**}p < 0.0005$. The error bars show the standard deviation of the scores.

When considering the sound changes of words, either CI coding strategies or levels of noises can affect sound changes of words which make the subjects identify incorrectly words. The findings for monosyllabic words demonstrate that sound changes can be divided into six main groups: consonant insertion, consonant deletion, consonant replacement, vowel changes, tone changes, and combined changes. Consonant insertion was found only in the final position, as in wâw (blame) → wâng (empty). On the other hand, consonant deletion and consonant replacement were found in both the initial and final positions, as in pla (fish) → la (donkey); kan (noun prefix) → ka (crow); tháe (genuine, real) → pháe (lose); chàp (catch) → chàt (arrange). Vowel changes were rarely found, as in bun (merit) → bon (above). Tone changes were found as well, as in má (horse) → mǎ (dog). Some changes may be combined together, for example lûk (child) → lung (uncle), which has two changes. One is a tone change from a falling tone to a mid tone. Another is a consonant replacement from /k/ to /ng/. Most sound changes caused confusion and misperception which resulted from the most consonant replacement, particularly in the initial position rather than in other positions, followed by a tone change. For bisyllabic words, either the preceding syllables or following syllables were replaced with the new syllables, caused by consonant replacement (initial/final/both positions), consonant deletion, tone changes or combined changes, such as lék-nǒi (little) → dék-nǒi (baby), ro-rót (wait for the bus) → ló-rót (wheel) and nâ-tàng (window) → nâ-ta (face). Notice that no matter what kinds of sound changes were found, their lexical meanings were also changed.

Generally, Thai syllable signals consist of three phoneme signals - the initial consonant, vowel, and final consonant (Theera-Umpon, Chansareewittaya, & Auephanwiriyakul, 2011). These signals are combined by temporal conditions. The tone is the F0 contours varied across the syllable. The vowel signals play an important role (Theera-Umpon et al., 2011) which are the most salient components and the important factors for identifying the number of syllables in each word. It is not surprising that the experiment found few confused vowels. In contrast, the experiment frequently found more confused consonants, especially consonant replacement. In addition, the initial consonant of words is more confusable than the final consonant of words. Listeners always recognize the initial consonants as the first phoneme. Thus, it is difficult to recognize either noisy words or processed words. Furthermore, the components of phonemes may be overlapping, for example, the components of an initial consonant may be combined with those of a vowel. Accordingly, if these components are distorted by any stage of signal processing (e.g. CI sound coding and adding noises), their sounds can be changed.

An example of the spectrograms with 22 channels (electrodes) for the CIS and the ACE strategies are shown in Fig. 3 for a Thai monosyllabic word "/yam/". The 12 out of 22 channels (electrodes) were selected for the ACE strategy. Then, the ACE has less speech information than the CIS but maintains the important information for speech understanding. The spectrograms of clean words and noisy words are compared as in Fig. 4 for the Thai monosyllabic word "/rue/" and the Thai bisyllabic word "/lék-nǒi /". The first row presents the spectrograms of the clean words. The second row presents the spectrograms of the monosyllabic word corrupted by SSN and the bisyllabic word corrupted by BBN at 5 dB SNR. It can be seen that the spectragrams of words were masked by noises which make the signals of words distorted from original signals. The noise signals may dominate the speech in some parts of signals, especially in speech signals with lower energy (e.g. consonants), such as the electrodes between 14 and 18 of the noisy monosyllabic word as in the bottom left of Fig. 4. It is not surprising that the listerners confuse and misperceive frequently consonants of words in noisy environments.

The current study was limited to the sets of tested words. It is difficult to deal with variation of related parameters for CI coding strategies (e.g. frequency scales and the number of selected channels), variation of related factors for complex listening environments (e.g. noise types and noise levels) and quantitative analysis of confusion (e.g. consonants, vowels and tone levels). Moreover, this study is unclear whether both the CIS and the ACE strategies arrange sufficiently for important information to encode tonal cues for the Thai language (Liu, Peng, Zhao, & Ni, 2017). The experiments with the diverse speech materials will highly support the investigation and evaluation of the effect of parametric variations of CI coding strategies to the performance of Thai speech intelligibility. However, this experiment is a preliminary analysis which can lead to develop new CI coding strategies for Thai speech intelligibility in the future.

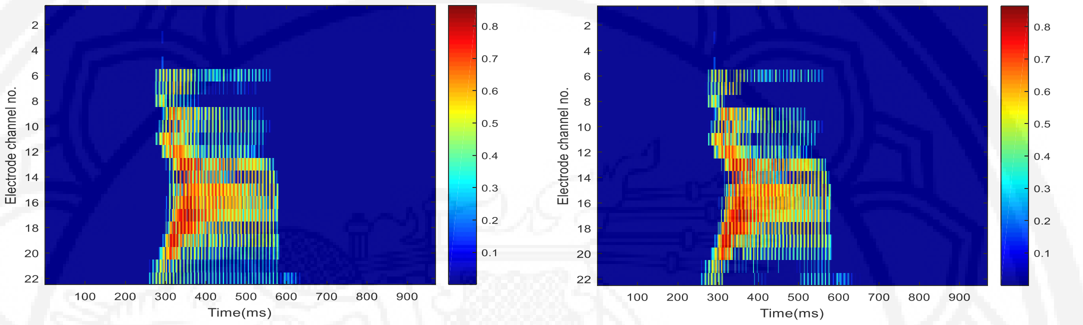
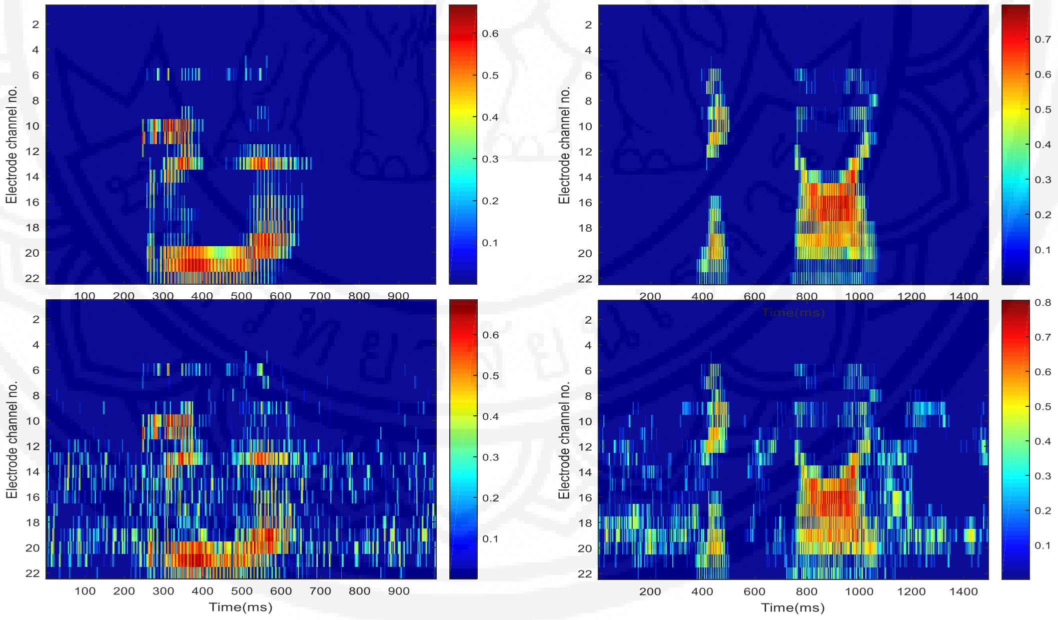


Figure 3 Spectrograms of the Thai monosyllabic word "/yam/" (guard) for the CIS (left) and the ACE (right) strategies. The electrode no. 1 and 22 represent the channels of the highest and the lowest frequency components, respectively.



(a) Monosyllabic word

(b) Bisyllabic word

Figure 4 An example of spectrograms for the ACE strategy. The spectrograms of the Thai monosyllabic word "/rue/" (ship) in the left column and the Thai bisyllabic word "/lék-noǐ /" (little) in the right column. The spectrograms of the clean words are presented in first row. The spectrograms of the noisy words at 5 dB SNR of speech-shaped noise for the monosyllabic word and 5 dB SNR of babble noise for the bisyllabic word, are presented in the second row.

Conclusion and Suggestions

In the present study about the performance of Thai speech intelligibility, two sound coding strategies, namely the CIS and the ACE strategies, were evaluated by twelve normal-hearing listeners using CI simulation based on a noise-band vocoder. The experimental results showed that speech intelligibility performance of the ACE was better than that of the CIS in all tested conditions. When compared to the CIS, the ACE had a significant performance in some conditions, particularly in the speech-shaped noise of 5 and 10 dB SNR for monosyllabic words and the speech-shaped noise of 0 dB SNR for bisyllabic words. The future research will be extended to evaluate performance of speech intelligibility in terms of more tested materials, noise types, SNR levels, and CI coding strategies, in order to cover daily-life noisy environments and find optimized CI coding strategies for individual CI users.

Acknowledgements

This research was supported by grants from Naresuan University. The authors would like to thank all the subjects who participated in this study. We would also like to thank Chadarthan Luangsawang, Siraprabha Kaewsri, and Patamawadee Doungta from the Department of Otolaryngology, Naresuan University for substantial help in collecting the data and useful suggestions and materials. Furthermore, we would like to thank anonymous reviewers for all comments and suggestions.

References

Chen, F., Hu, Y., & Yuan, M. (2015). Evaluation of Noise Reduction Methods for Sentence Recognition by Mandarin-Speaking Cochlear Implant Listeners. *Ear and Hearing, 36*(1), 61-71.

Chen, F., & Loizou, P. C. (2011). Predicting the Intelligibility of Vocoded Speech. *Ear and Hearing, 32*(3), 331-338.

Cochlear. (2002). *ACE™ and CIS DSP Strategies Software Requirements Specification,Part Number: N95287F Issue 1*. Lane Cove, New South Wales: Australia.

Fan-Gang, Z., Rebscher, S., Harrison, W., Xiaoan, S., & Haihong, F. (2008). Cochlear implants: system design, integration, and evaluation. *IEEE Reviews in Biomedical Engineering*, *1*(1), 115-142.

Kangsanarak, B. (1980). *Development of Thai Spondee Words for Speech Audiometry.* Mahidol University, Bangkok.

Kasemsuwan, L., Cheewaruangroj, W., Tungkeeratichai, J., Bhongmakapat, T., Thawin, C., Lertsukprasert, K., . . . Laothamatas, J. (2010). Fourteen-years experience with cochlear implantation in Ramathibodi Hospital. *Journal of the Medical Association of Thailand, 93*(12), 1399-1405.

Kiefer, J., Hohl, S., Sturzebecher, E., Pfennigdorff, T., & Gstoettner, W. (2001). Comparison of speech recognition with different speech coding strategies (SPEAK, CIS, and ACE) and their relationship to telemetric measures of compound action potentials in the nucleus CI 24M cochlear implant system. *Audiology, 40*(1), 32-42.

Komalarajun, S. (1979). *Development of Thai Speech Discrimination Materials.* Mahidol University, Bangkok.

Lin, Y.-H. (2013). *The Sounds of Chinese*. Cambridge: Cambridge University Press.

Liu, H., Peng, X., Zhao, Y., & Ni, X. (2017). The effectiveness of sound-processing strategies on tonal language cochlear implant users: A systematic review. *Pediatric Investigation*, *1*, 35-42.

Loizou, P. C. (2006). Speech processing in vocoder-centric cochlear implants. In A. R. Moller (Ed.), *Cochlear and Brainstem Implants*, *13 July 2006* (pp. 109-143). Basel, Switzerland: Karger.

Nogueira, W., Buchner, A., Lenarz, T., & Edler, B. (2005). A psychoacoustic "NofM"-type speech coding strategy for cochlear implants. *Eurasip Journal on Applied Signal Processing*, *2005*(18), 3044-3059.

Qin, M. K., & Oxenham, A. J. (2003). Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers. *Journal of the Acoustical Society of America*, *114*(1), 446-454.

Skinner, M. W., Holden, L. K., Whitford, L. A., Plant, K. L., Psarros, C., & Holden, T. A. (2002). Speech recognition with the nucleus 24 SPEAK, ACE, and CIS speech coding strategies in newly implanted adults. *Ear and Hearing*, *23*(3), 207-223.

Theera-Umpon, N., Chansareewittaya, S., & Auephanwiriyakul, S. (2011). Phoneme and tonal accent recognition for Thai speech. *Expert Systems with Applications*, *38*(10), 13254-13259.

Vaewvichit, K., & Luangpitakchumpol, P. (1999). Cochlear implantation in Thailand. *The Journal of Laryngology & Otology*, *113*(6), 515-517.

Wilson, B. S. (2017). The cochlear implant and possibilities for narrowing the remaining gaps between prosthetic and normal hearing. *World J Otorhinolaryngol Head Neck Surg*, *3*(4), 200-210.

Wilson, B. S., & Dorman, M. F. (2008). Cochlear implants: A remarkable past and a brilliant future. *Hear Res*, *242*(1-2), 3-21.

Wilson, B. S., & Dorman, M. F. (2008). Cochlear implants: Current designs and future possibilities. *Journal of Rehabilitation Research and Development*, *45*(5), 695-730.

Wouters, J., McDermott, H. J., & Francart, T. (2015). Sound coding in cochlear implants: from electric pulse to hearing. *IEEE Signal Processing Magazine*, *32*(2), 67-80.

Xu, L., & Zhou, N. (2011). Tonal Languages and Cochlear Implants. *Auditory Prostheses*, *39*, 341-364.