Speech intelligibility enhancement for Thai-speaking cochlear implant listeners

Siriporn Dachasilaruk[1], Niphat Jantharamin[2], Apichai Rungruang[3]
[1,2]Department of Electrical and Computer Engineering, Faculty of Engineering, Naresuan University, Thailand
[3]Department of English Language, Faculty of Humanities, Naresuan University, Thailand

Article Info

Article history:

Received Aug 03, 2018
Revised Nov 23, 2018
Accepted Dec 12, 2018

Keywords:

Sound coding
Spectral subtraction
Speech enhancement
Speech intelligibility
Weiner filter

ABSTRACT

Cochlear implant (CI) listeners encounter difficulties in communicating with other persons in noisy listening environments. However, most CI research has been carried out using the English language. In this study, single-channel speech enhancement (SE) strategies as a pre-processing approach for the CI system were investigated in terms of Thai speech intelligibility improvement. Two SE algorithms, namely multi-band spectral subtraction (MBSS) and Weiner filter (WF) algorithms, were evaluated. Speech signals consisting of monosyllabic and bisyllabic Thai words were degraded by speech-shaped noise and babble noise at SNR levels of 0, 5, and 10 dB. Then the noisy words were enhanced using SE algorithms. The enhanced words were fed into the CI system to synthesize vocoded speech. The vocoded speech was presented to twenty normal-hearing listeners. The results indicated that speech intelligibility was marginally improved by the MBSS algorithm and significantly improved by the WF algorithm in some conditions. The enhanced bisyllabic words showed a noticeably higher intelligibility improvement than the enhanced monosyllabic words in all conditions, particularly in speech-shaped noise. Such outcomes may be beneficial to Thai-speaking CI listeners.

Corresponding Author:

Siriporn Dachasilaruk,
Department of Electrical and Computer Engineering,
Faculty of Engineering,
Naresuan University, Phitsanulok 65000, Thailand.
Email: siripornd@nu.ac.th, dsiriporn@hotmail.com

1. INTRODUCTION

Most CI listeners can achieve high speech intelligibility which is close to the capability of normal-hearing (NH) listeners in quiet listening environments. This is because almost all the sound coding strategies used by modern CI devices perform well in quiet listening environments [1]. However, most CI listeners suffer from decreased speech intelligibility more than NH listeners in noisy listening environments. The higher the noise level, the lower the speech intelligibility performance [2]. One of the specific limitations of CI devices in terms of frequency, temporal and amplitude resolutions [3] is related to transmitting speech information to the auditory nerves. Another limitation is the effect of channel interaction, which results from the overlap of electrical fields between electrodes [4]. Electric stimulation of one electrode may be distorted by the stimulation of other electrodes. Such interactions can decrease intelligibility performance. Therefore, CI researchers have increasingly attempted to improve speech intelligibility performance, particularly by developing speech enhancement (SE) strategies for use in adverse noisy environments.

Generally, single-channel SE strategies are used in most traditional CI systems, and this can be extended to apply to multi-channel SE strategies. Therefore, single-channel SE strategies were employed in this study. Several studies have indicated that these single-channel SE algorithms improved speech intelligibility significantly for hearing-impaired (HI) listeners. SE algorithms such as the pre-processing approach have been applied to CI systems, including subspace-based, Weiner filter, and spectral subtractive algorithms. Loizou et al. [5] demonstrated that the subspace-based algorithm proposed by Hu and Loizou [6] significantly improved sentence recognition in speech-shaped noise at a 5 dB signal-to-noise ratio (SNR) among a group of fourteen Clarion CI listeners, with an average improvement of 44%. However, this algorithm can also provide recognition benefits for stationary noise (e.g. speech-shaped noise (SSN)), but the algorithm does not guarantee improvement for non-stationary noises (e.g. babble noise (BBN)). Bolner et al. [7] showed that a Weiner filter (WF) based on a priori SNR estimates [8] significantly improved sentence recognition in a SSN condition at 0 dB SNR in a group of ten NH listeners, but there was no improvement in a BBN condition. Additionally, a study by Koning et al. [9] showed that the WF algorithm, applied as an envelope-weighting approach, provided both speech intelligibility and speech quality improvement for a group of six Dutch-speaking CI listeners and six Dutch-speaking NH listeners.

Over almost four decades, spectral subtraction (SS) algorithms have been developed in many versions, and some of these have been applied in CI systems. An SS algorithm referred to as the INTEL SE algorithm was first applied by Hochberg et al. [10]. Consonant-vowel-consonant words corrupted by SSN at SNR ranging from -10 to 25 dB for NH listeners and from -5 to 25 dB for CI listeners were processed by the INTEL SE algorithms. The enhanced words were presented to ten NH listeners and ten Nucleus CI listeners. Word recognition was significantly improved for CI listeners but not for NH listeners. A study by Weiss [11] indicated that when the noisy speech signals were enhanced by the INTEL SE algorithm, the error of the second formant extraction was considerably reduced in the Nucleus implant coding strategy. These effects were used to improve speech perception in the prior study.

Yang and Fu [12] found significant improvements with the SS algorithm proposed by Gustafsson et al. [13] when it was applied to sentence recognition in SSN at different SNRs (i.e. 0, 3, 6, and 9 dB) in a group of seven CI listeners who used different CI devices (i.e. Nucleus, Med-El, and Clarion). Verschuur et al. [14] indicated that sentence recognition with the nonlinear SS (NSS) algorithm proposed by Lockwood and Boudy [15] was significantly improved in SSN at both 5 and 10 dB SNRs for seventeen Nucleus CI listeners. However, such benefits may be limited to suppressing non-stationary noise. A later study by Kallel et al. [16] applied the NSS algorithm proposed by Berouti et al. [17] and the multi-band SS (MBSS) algorithm proposed by Kamath and Loizou [18] to three bilateral Neurelec CI listeners and fifty NH listeners. The results showed that the average word recognition improvement was 4–9% for bilateral Neurelec CI listeners and 7–13% for NH listeners at all SNRs (i.e. -3, 0, 3, and 6 dB). Moreover, the results also indicated that the MBSS algorithm enhanced speech intelligibility more than the NSS algorithm for single and multiple interfering noise sources.

Nevertheless, most SE strategies for CI listeners have been evaluated using the English language. A few studies have evaluated strategies using French, Hebrew, Dutch/Flemish, and Chinese, but SE strategies have never been evaluated using the Thai language. Different languages have different acoustic cues and phonemics, which may produce different intelligibility performances using the same SE techniques. English is a non-tonal language, whereas Thai is a tonal language which is similar to many Asian languages (e.g. Chinese and Vietnamese). A tonal language uses a tonal level which is distinguished by the fundamental frequency (F0) contours to represent lexical meaning. Each tone of a word represents a different meaning. Normally a Thai syllable consists of an initial consonant (a single/clustered consonant), a vowel (short/long), an optional final consonant and a tonal level. There are five distinctive tones in Thai syllables: the middle /¯/, the low /`/, the falling /ˆ/, the high /´/, and the rising /˘/. The Thai tones of monosyllabic words are commonly available. Examples of monosyllabic words with five tones differentiating their meaning are /pāa/ (throw), /pàa/ (forest), /pâa/ (aunt), /páa/ (dad) and /pǎa/ (dad). These components are important to the performance of speech intelligibility among Thai-speaking CI listeners.

No studies have specifically evaluated SE methods with Thai-speaking CI listeners. Therefore, the objective of the present study is to investigate the speech intelligibility performance of existing SE algorithms and to assess whether different SE algorithms provide different intelligibility performance in various noisy environments for Thai-speaking CI listeners. The investigation of the improvement of Thai word recognition concentrates on SE algorithms as the pre-processing approach, namely multi-band spectral subtraction (MBSS) and Weiner filter (WF) algorithms. Both achieve a trade-off between effective noise reduction, speech distortion, and low computation costs for real-time implementations [8, 14]. The WF algorithm with nonvocoded speech has shown high speech intelligibility scores in three languages: English, Chinese, and Japanese [19]. The MBSS algorithm is one version of the spectral subtraction algorithm which has been shown to yield a speech intelligibility improvement for French-speaking and Mandarin-speaking CI listeners. Some previous studies support selecting both SE algorithms. A speech intelligibility evaluation was conducted on NH listeners in a CI simulation with a noise-band vocoder. The present study was extended from a study by Dachasilaruk et al. [20] with a larger number of subjects.

2. SPEECH ENHANCEMENT FOR THE CI SYSTEM

2.1. Speech Enhancement Algorithms

Figure 1 presents a CI simulation with a noise-band vocoder based on speech enhancement. The noisy speech is processed with a SE algorithm to generate enhanced speech. After that, the enhanced speech is fed into the CI system to produce vocoded speech. Further descriptions of the MBSS and WF algorithms are given in Kamath and Loizou [18] and Scalart and Vieira [8] respectively. Both algorithms are briefly described in this section. Assume that noisy speech signals (*y*) at a sampling rate of 16 kHz are generated by adding noise (*n*) to clean speech signals (*x*). Then, the power spectrum of the noisy speech signals can be approximately estimated as follows:

$$|Y(k)|^2 \approx |X(k)|^2 + |N(k)|^2 \tag{1}$$

The MBSS algorithm is slightly different to the NSS algorithms. The MBSS uses a factor of subtraction estimated in each frequency bin and each frequency band, whereas the NSS uses this factor estimated in each frequency bin. The concept of the MBSS is that the characteristics of the noise spectrum may not affect the speech spectrum equally across the entire frequency band. The noise spectrum may affect some frequency bands more or less than others. Therefore, spectral subtraction is performed separately in each frequency band. In the i^{th} band, the spectrum of the clean speech is estimated as:

$$\left|\hat{X}_i(k)\right|^2 = |Y_i(k)|^2 + \alpha_i \delta_i \left|\hat{N}_i(k)\right|^2, b_i \le k \le e_i \tag{2}$$

where $\left|\hat{N}(k)\right|^2$ is the estimated spectrum of the noise signal, b_i and e_i are the start and stop bins of the i^{th} band, and α_i and δ_i are the over-subtraction and weight factors of the i^{th} band, respectively. The weight factor can be individually set for each band.

For the WF algorithm, the gain function *g(k)* is expressed with a *priori* SNR ξ_k. The ξ_k is estimated from the past and present estimates of ξ_k as a weighted combination. The $\hat{\xi}_k(m)$ of the present frame *m* is estimated as follows:

$$g(k) = \frac{\xi_k}{\xi_k + 1} \tag{3}$$

$$\hat{\xi}_k(m) = \alpha \frac{\left|\hat{X}_k(m-1)\right|^2}{|N_k(m-1)|^2} + (1-\alpha)\max\left(\frac{|Y_k(m)|^2}{|N_k(m)|^2} - 1, 0\right) \tag{4}$$

where *α* denotes a smoothing constant (*α*=0.98), and $\hat{X}_k(m-1), Y_k(m)$ and $N_k(m)$ are the spectrum of the enhanced speech signal at the past frame *m*-1, the spectrum of the noisy speech and the noise signal at the present frame *m*, respectively.

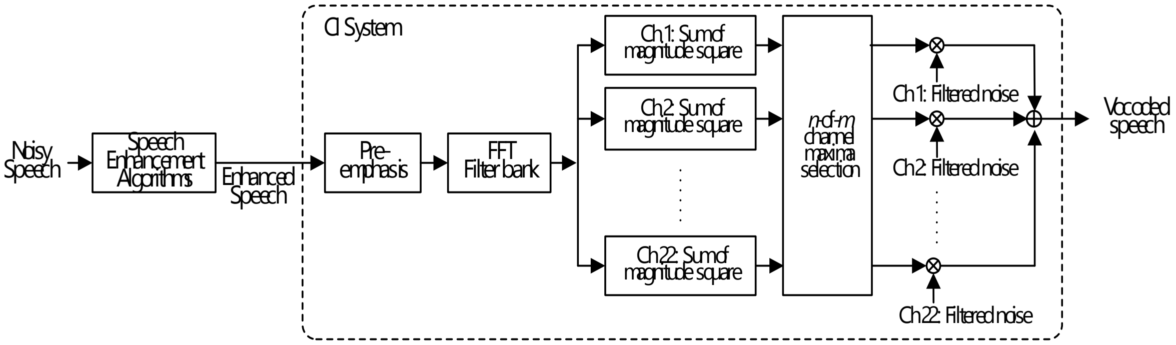

Noisy Speech
Speech Enhancement Algorithms
Enhanced Speech
CI System
Pre-emphasis
FFT Filter bank
Ch1: Sum of magnitude square
Ch2: Sum of magnitude square
Ch22: Sum of magnitude square
n-of-m channel maxima selection
Ch1: Filtered noise
Ch2: Filtered noise
Ch22: Filtered noise
Vocoded speech

Figure 1. Block diagram of a CI simulation based on speech enhancement

2.2. Sound Coding Strategies

Commercial manufacturers of CI devices propose many sound coding strategies such as the Continuous Interleaved Sampling (CIS) strategy and the Advanced Combination Encoder (ACE) strategy. Generally, the CIS strategy is proposed in CI devices made by all manufacturers, and it has different implementations depending on the manufacturer. The Cochlear Company, producing Nucleus CI devices, offers both the CIS and ACE strategies. The difference between these strategies lies in the channel maxima selection stage. Channel maxima selection is performed in the ACE but not in the CIS. A few studies have revealed that most Nucleus CI listeners preferred the ACE over the CIS, and the mean scores using the ACE show significantly higher speech intelligibility than those using the CIS [21], [22]. Moreover, the preferred strategy corresponded with the speech intelligibility outcome.

The ACE strategy can be described as an *n*-of-*m* strategy [23]. The speech signal is decomposed into *m* channels related to the number of electrodes, but only the *n* channels with maximum amplitudes are selected for simultaneous stimulation. The concept of the ACE strategy is to increase temporal resolution and reduce redundant information in speech. The most important channels containing important speech information can be updated more frequently by removing the less significant channels [24]. This strategy may reduce the overall SNR level and presumably reduces channel interaction [25]. Additionally, the power consumption of electrical stimulation can be decreased, and this may lengthen battery life for CI devices [26].

As shown in Figure 1, in part of the CI system the enhanced speech was filtered by a pre-emhasis filter to amplify the high-frequency components of speech information. Then, the frame-by-frame processing of 128 samples with an overlap of 75% was applied to the pre-emphasized speech. The greater the overlap of the frame, the higher the channel stimulation rates. Each frame of the pre-emphasized speech was decomposed using the fast-Fourier transform (FFT) into uniform frequency bands (128 bins), with the frequency band of each bin at 125 Hz. Only the first 64 bins were used, to generate a frequency resolution of 22 channels. The powers of consecutive bins were summed within frequency ranges specified in the CI system. The cutoff frequencies of the 22 channels were 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. After that, the 12 envelope channels with the largest amplitudes were modulated by white noise with the same cutoff frequencies as the FFT filter bank. Finally, vocoded speech was synthesized by summing all the selected channels of the modulated signal. The vocoded speech was then presented to NH listeners for testing.

3. PERFORMANCE EVALUATION

All the Thai words in this study were from a corpus which is commonly used in clinical practice with HI listeners. In this study the Thai word test was divided into 8 lists of monosyllabic words and 8 lists of bisyllabic words [27], [28]. Each list had 25 words and the total words are 400 words. After the words were selected from the corpus, all the recorded words were corrupted by speech-shaped noise (SSN) at SNR levels of 0 and 5 dB, and babble noise (BBN) at SNR levels of 5 and 10 dB. The levels of SNR were carefully chosen to avoid floor and ceiling effects [7], [29], as well as particularly noisy and enhanced monosyllabic words. Then, the noisy words were processed using the MBSS and WF algorithms. The enhanced and noisy words were processed using the ACE strategy to produce the vocoded speech signals. The vocoded speech signals were presented to NH subjects in a total of 24 conditions ([2 SE algorithms + 1 unprocessed SE algorithm] × 2 word types × 2 noise types × 2 SNR levels).

Twenty NH subjects (14 males, 6 females, age range from 20 to 40 and mean age 26) participated in this experiment. All subjects were native speakers of Thai, and were undergraduate students and staff at a Thai public university. Otoscopy was undertaken for all subjects to check for any abnormalities in their middle ears. Then, all subjects undertook a pure tone audiogram test to confirm that they had NH thresholds (< 25 dB HL, between 0.25 and 8 kHz). All subjects were paid for their participation and they all signed a consent form. This experiment was approved by the Ethics Committee of the university.

The vocoded speech signals were presented using a laptop, unilaterally through a headphone. The subjects used only one preferred ear, the one that was most comfortable for them, to listen to the vocoded speech signals in all tested conditions. The volumes of the vocoded speech signals were calibrated to be at a comfortable conversional level. Each subject was assessed in a total of 24 conditions over two sessions on separate days (12 conditions per session, one session per day), with a break of at least one week between the two sessions to avoid learning effects. Testing lasted approximately one hour in each session.

Before the actual tests were carried out, the researchers offered training tests to ensure that the subjects clearly understood how to do the tests. In the training tests the subjects were asked to listen to both noisy and enhanced words in all conditions during a 5-minute test, to familiarise themselves with the vocoded speech signals. They were trained in both sessions before they undertook the actual tests. In the actual tests, the subjects listened to the words and wrote them down on their papers. They could guess the words if they were uncertain. Subjects were scored based on how many words they identified correctly in each list. One word list was administered per condition, and each list contained one hundred words. No word list was repeated across the conditions in each session for each subject. The list-to-condition mapping and the order of tested conditions in each session were randomized for each subject. The subjects did not know which condition would be tested and what the tested conditions would be. In order to avoid listening fatigue which may affect performance, the subjects were given a break every 20 minutes during the actual test, or whenever they required a rest.

The scores of correct words were averaged based on the percentage of words identified correctly. Then the scores were statistically analyzed using SPSS software. An analysis of variance (ANOVA) with repeated measures was carried out to investigate the differences between mean scores in terms of SE algorithms, SNR levels, noise types, and word types. A post hoc Bonferroni-corrected test with a multiple paired comparison was employed to examine the individual relationships between the mean scores in each condition.

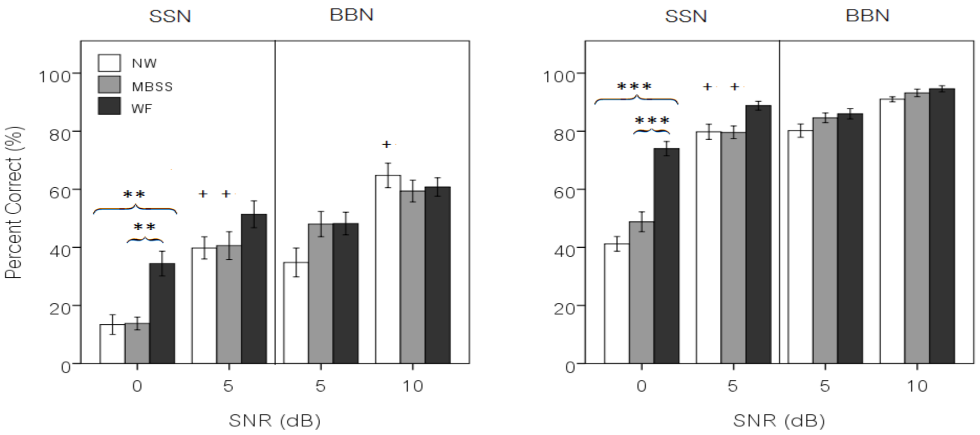
4. RESULTS AND DISCUSSION

The mean percentage correct scores of 20 NH subjects in 24 conditions are shown in Figure 2. All the conditions consisted of noisy words and enhanced words due to different SE algorithms, SNR levels, noise types, and word types. The results of noisy and enhanced words showed considerably increased mean scores as SNR levels increased at the same noise type and word type. For the noisy words, the mean scores of the bisyllabic words with BBN increased slightly as SNR levels increased. The mean scores showed extremely high intelligibility at 5 and 10 dB SNR. At 5 dB SNR the mean scores for noisy words with SSN were slightly higher than those for noisy words with BBN for monosyllabic words, and almost the same for bisyllabic words. The enhanced words revealed higher mean scores than the noisy words in most conditions. The WF reflected considerably better performance improvement than the MBSS in almost all conditions, except for monosyllabic words at 5 dB SNR of BBN.

For the monosyllabic words, the WF reflected a greater intelligibility improvement than the MBSS, especially for the condition of SSN at 0 and 5 dB SNR. Both algorithms yielded almost the same mean scores for BBN at 5 and 10 dB SNR. However, the enhanced words with both algorithms showed lower mean scores than the noisy words for BBN at 10 dB SNR. For the bisyllabic words, the WF algorithm showed a noticeably higher improvement than the MBSS, especially in the condition of SSN at 0 and 5 dB SNR. Both algorithms illustrated a slight improvement for BBN. The overall mean scores for the bisyllabic words were higher than those for the monosyllabic words.

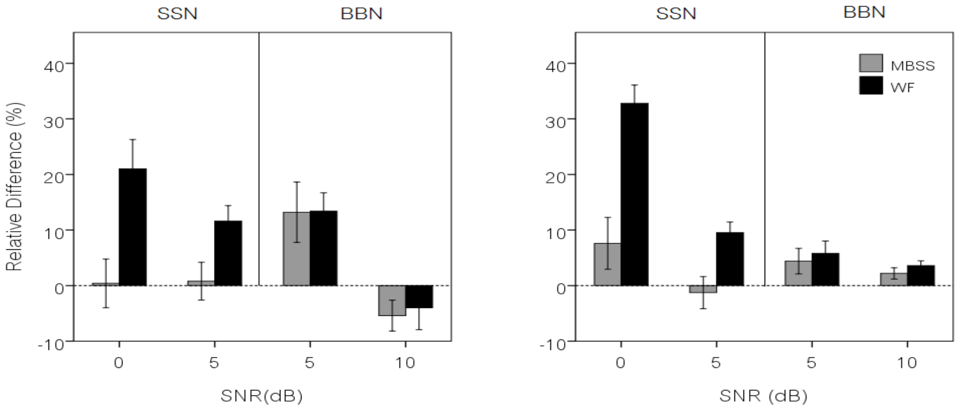
A two-way ANOVA with repeated measures was used to explore the two factors of SE algorithm and SNR level. For the monosyllabic words with SSN, the statistical analysis revealed a significant effect of SE algorithm [$F(2,38)=17.60$, $p<0.0005$] and SNR level [$F(1,19)=68.73$, $p<0.0005$]. Post hoc tests for SE algorithms indicated that the WF produced significantly higher intelligibility scores than the noisy words, and showed significantly better performance than the MBSS at 0 dB SNR. Post hoc tests of SNR levels revealed that an increased SNR level provided significantly improved intelligibility scores for noisy words and the MBSS ($p<0.0005$), but this difference was not statistically significant for the WF. For the monosyllabic words with BBN, there were significant effects of SNR level [$F(1,19)=105.73$, $p<0.0005$] and a significant interaction effect between SE algorithm and SNR level [$F(2,38)=7.95$, $p<0.05$]. Post hoc tests of SNR levels indicated that the noisy words at 10 dB SNR showed significantly higher intelligibility scores than those at 5 dB SNR. For the bisyllabic words with SSN, the statistical analysis indicated a significant effect of SE algorithm [$F(2,38)=50.28$, $p<0.0005$], a significant effect of SNR level [$F(1,19)=268.57$, $p<0.0005$] and a significant interaction effect between SE algorithm and SNR level [$F(2,38)=12.27$, $p<0.0005$]. Post hoc tests of SE algorithms and SNR levels indicated that the multiple paired comparison yielded the same results as the monosyllabic words with SSN.

The relative difference of mean scores between noisy words and enhanced words is shown in Figure 3. There was a considerable increase in intelligibility improvement for enhanced words with SE algorithms for both word types which were evaluated in the same tested conditions. Except for monosyllabic words in BBN at 10 dB SNR, intelligibility performance decreased. The WF exhibited a high improvement but the MBSS showed a low improvement in both word types. In terms of the overall result of intelligibility improvement in the whole condition, SE algorithms improved by approximately 3% for the MBSS and by 12% for the WF. A trend of decreased intelligibility performance with increased SNR levels was found in both SE algorithms. The WF yielded better intelligibility than the MBSS at the same SNR levels.



(a) Monosyllabic words (b) Bisyllabic words

Figure 2. The mean percentage correct scores of 20 normal-hearing listeners for the noisy words (NW) and the enhanced words with the MBSS and WF algorithms are shown for speech-shaped noise (SSN) at 0, 5 dB SNR and babble noise (BBN) at 5 and 10 dB SNR. The plus (+) denotes that the mean percentage correct scores are significantly higher than those at the lower SNR level (p<0.0005). An asterisk (*) denotes that the mean percentage correct scores are significantly higher than those at the same SNR level; **p<0.05, ***p< 0.0005. The error bars indicate the standard deviation of the scores



(a) Monosyllabic words (b) Bisyllabic words

Figure 3. Relative differences in intelligibility scores between noisy and enhanced words with the MBSS and WF algorithms are shown for speech-shaped noise (SSN) at 0, 5 dB SNR, and for babble noise (BBN) at 5, 10 dB SNR. Positive numbers indicate an increased intelligibility performance whereas negative numbers represent a decreased intelligibility performance. The error bars refer to the standard deviation of the scores

Examples of the electrodograms for of the monosyllabic word “/yāam/” and the bisyllabic word “/nâe-nōn/” are presented in Figure 4. As can be clearly seen from the electrodograms, the MBSS and WF algorithms can reduce noise. However, the content of speech information was noticeably reduced for the enhanced words with the MBSS, and residual noise still remained in some segments of the enhanced words with the WF. The WF preserved more speech information at low frequencies ranging from 187 to 563 Hz or in the electrode channels between 20 and 22. Such preservation of information by the WF in noisy conditions echoed a significant intelligibility improvement.

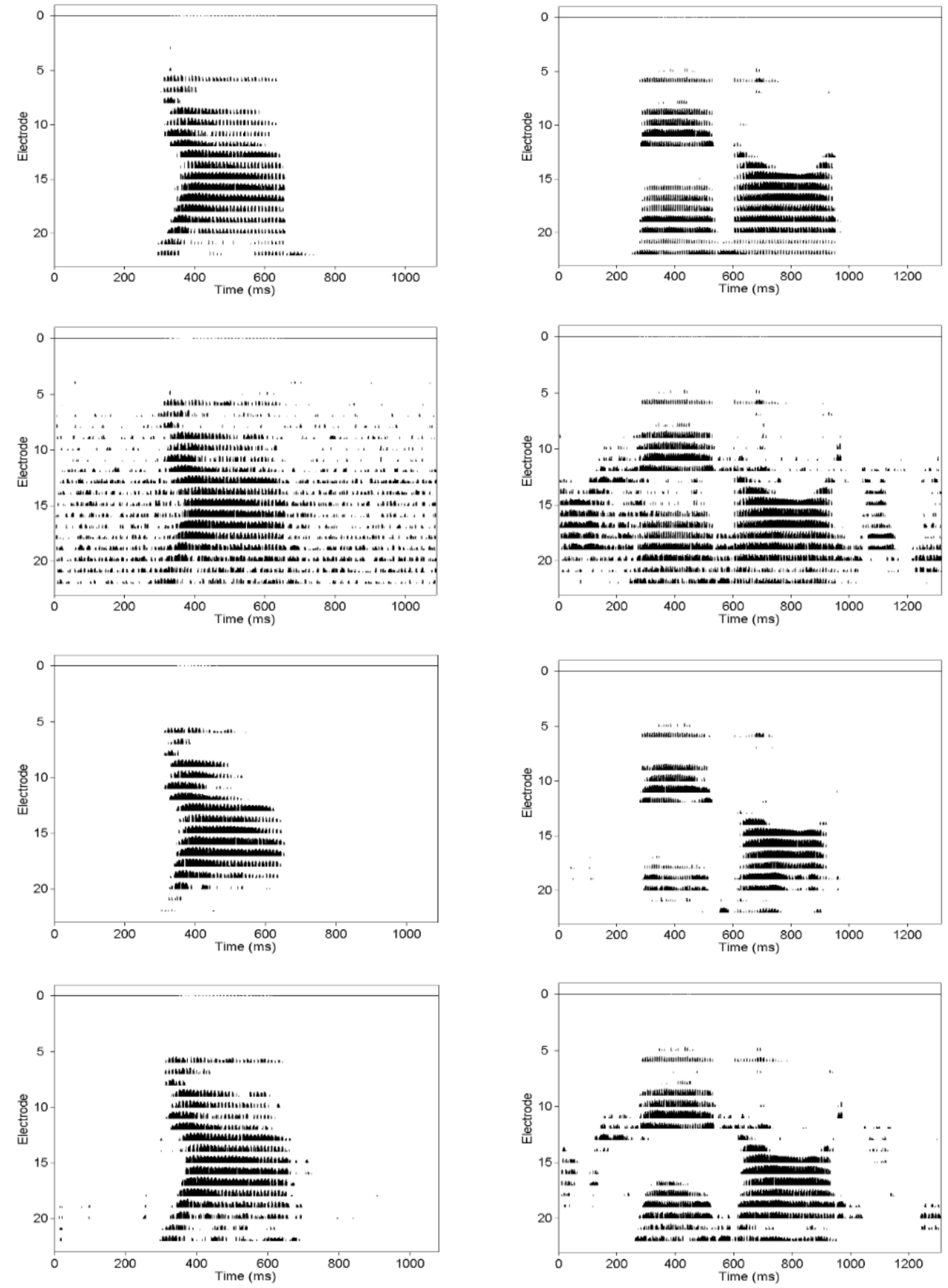


Figure 4. Electrodograms of the monosyllabic word "/yāam/" (guard) in the left column and the bisyllabic word "/nâe-nōn/" (certain) in the right column. The first row shows the clean words. The second row represents the noisy words at 5 dB SNR speech-shaped noise for the monosyllabic word and 5 dB SNR babble noise for the bisyllabic word. The third and fourth rows show the enhanced words with the MBSS and WF algorithms, respectively

The outcomes of the present study have demonstrated that there is potential for single-channel SE algorithms (i.e. MBSS and WF) under various noisy conditions to be able to improve intelligibility performance for Thai CI listeners. Some of the results of the present study were consistent with those of Chen et al. [29]. This study found that the best intelligibility performance was achieved by the WF for both stationary noise (e.g. SSN) and non-stationary noise (e.g. BBN). Because of the fact that the WF performs a trade-off between effective noise reduction and speech distortion [30], the WF enhanced speech may contain more residual noise but also more preserved lexical information. The majority of the preserved lexical information has low-frequency components, which are important information for speech understanding [4]. Not surprisingly, the WF enhanced speech exhibits higher word recognition scores. For the MBSS, the noise spectrum naturally affects the speech spectrum across various frequency bands. The noise spectrum may be over-estimated and over-subtracted from the spectrum of noisy speech in each frequency band. Some segments of lexical information may be aggressively removed or distorted by processing from the MBSS. The residual information in enhanced speech is insufficient to understand the lexical meaning. Thus, word recognition for the MBSS enhanced speech is almost equal to or worse than that of noisy speech in some conditions.

Bisyllabic words reveal significantly higher intelligibility scores than monosyllabic words in all tested conditions. Generally, the bisyllabic words had longer speech durations and contained more lexical information than the monosyllabic words. When they were masked by noise or some segments of lexical information were removed or degraded by the SE algorithms and CI coding strategy, the listeners were able to understand the content of the words and guess the bisyllabic words better than the monosyllabic words.

Disguised words which result from noise, SE algorithms and CI coding strategies change sounds and affect intelligibility. The sound changes of monosyllabic words can be classified into six major types. These are consonant insertion, consonant replacement, consonant deletion, vowel changes, tone changes, and others. Consonant insertion was found only in the final position, as in sĭe (spoil) → sĭeng (sound). However, consonant replacement and consonant deletion were found in both the initial and final positions, as in tam (follow) → ta (eye); pla (fish) → la (donkey); hǒm (fragrance, onion) → phǒm (thin); fan (tooth) → fang (listen). Vowel changes were rarely found, as in dang (famous) → dam (black). Tone changes were found as well, as in klông (camera) → klong (drum). Interestingly, some changes cannot be catagorized. For example, ìm (have enough food) → kin (eat) shows that an initial consonant was inserted and a consonant replancement was made (/n/ → /m/). Another example is mâk (many) → ma (come), which also reveals two changes. One was a tone change from a falling tone to a mid tone. Another was a consonant deletion; /k/ was deleted.

For bisyllabic words, the replacement of preceding syllables and following syllables results in sound changes. Either preceding syllables or following syllables were replaced by new syllables which reflect consonant replacement and tone changes, such as ro-rót (wait for the bus) → ló-rót (wheel); fai-fá (electricity) → fai-pà (forest fire); lék-nǒi (little) → dék-nǒi (baby). Another change in the following syllables was a consonant deletion and a tone change, as in nâ-tàng (window) → nâ-ta (face). Notice that no matter what types of sound changes were found, the sound changes led to new meanings.

In principle, Thai syllable signals are composed of the initial consonant, vowel and final consonant signals [31]. The initial consonant and vowel signals contain either low or high frequency components. The final consonant signals contain only low frequency components. The tone is the change of F0 contours across the syllable. The signals for the three phonemes are combined by temporal conditions, and the vowel signal acts as the main syllable signal [31]. Therefore, the vowel signal is the largest component and always dominates the others. In other words, when compared to the vowels the initial and final consonants play minimal roles.

The findings rarely found confusable vowels. This is not surprising since vowels, in general, are the most salient components and they are the key factors for identifying the number of syllables in each word. On the contrary, the study frequently found confusable consonants, especially consonant replacement. A consonant in the initial position is more confusable than a conconant in the final position. This is not surprising; initial consonants are always the first phoneme, and listeners need to perceive them earlier than other phonemes. Thus, it is not easy to recognize these sounds, either in processed speech or noisy speech. Moreover, the components of phonemes may be overlapping. For instance, the components of an initial consonant and a vowel may be combined. Consequently, if these components are distorted by noise or any processing stage, this affects the sound and causes the meaning to change.

In the present study, the sets of tested words had some limitations when dealing with related factor variations according to SE algorithms or noisy environments, as in the studies by Li et al. [19] and Chen et al. [29]. The present corpus has been employed in audiology clinics for over forty years, particularly with HI listeners. As a result, for the Thai language corpus, existing speech materials are inadequate and not very practical for either clinics or research studies. This diversity in speech materials is an important issue. The test materials of consonants, vowels, tones or words are more suitable for analyzing speech information to represent a HI listener’s perception ability and reveal informative results about the effects of factors or parametric variation. The sentence test materials involved many factors (e.g. meaning, context, rhythm, etc.) and may be more appropriate for real-life communication. Thus, the development of Thai speech materials should be undertaken to support more effective evaluations and treatments. In other words, a large corpus of diverse speech materials is needed, which will benefit audiological evaluations to investigate speech intelligibility not only among HI listeners but also for CI listeners. Nonetheless, even with these limitations in relation to the corpus, the present study is a stepping stone to future studies with CI listeners.

There remains an enormous gap in knowledge and understanding of Thai speech intelligibility among Thai-speaking CI listeners. As a result, the intelligibility of Thai speech sounds should be intensively investigated in terms of Thai language features (e.g. consonants, vowels, and tones), speech coding strategies, SE algorithms, objective and subjective measurements, adverse environments (e.g. noise types and SNR levels), and listeners (both NH and CI listeners). New SE algorithms should be developed using other techniques such as deep machine learning [32], compressive sensing [33], [34] and so on. These algorithms should be specifically adapted to the auditory perception of CI listeners to optimally improve performance of either speech intelligibility or quality. Iin turn, similarities or differences in perception patterns and cross-linguistic observations may then be properly applied in many systems and related areas for future research.

5. CONCLUSION

In the present study about Thai intelligibility performance, two single-channel SE algorithms, namely MBSS and WF algorithms, were evaluated by twenty NH listeners using CI simulation with a noise-band vocoder. The results revealed that speech intelligibility performance was improved for both SE algorithms in most tested conditions, and was statistically significantly improved by the WF algorithm in some conditions. The WF algorithm performed better than the MBSS algorithm in nearly all conditions. This trend in outcomes will be useful information for Thai-speaking CI listeners in future investigations of the speech intelligibility performance of existing SE algorithms, and in developing either new SE algorithms or new sound coding strategies for auditory prostheses in future studies.

ACKNOWLEDGEMENTS

This research was supported by grants from Naresuan University. The authors would like to thank all the subjects who participated in this study. Furthermore, 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 Dr. Rattinan Tiravanitchakul and Rada Dara from the Department of Communication Science and Disorders, Faculty of Medicine, Mahidol University, Thailand for their useful suggestions.

REFERENCES

[1] Kokkinakis K, Azimi B, Hu Y, Friedland DR. Single and Multiple Microphone Noise Reduction Strategies in Cochlear Implants. *Trends in Amplification*. 2012;16(2):102-16.

[2] Hu Y, Loizou PC, Li N, Kasturi K. Use of a sigmoidal-shaped function for noise attenuation in cochlear implants. *Journal of the Acoustical Society of America*. 2007;122(4):128-34.

[3] Qin MK, Oxenham AJ. Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers. *Journal of the Acoustical Society of America*. 2003;114(1):446-54.

[4] Loizou PC. Speech processing in vocoder-centric cochlear implants. In: Moller AR, editor. Cochlear and Brainstem Implants. 64. Basel, Switzerland: Karger; 2006. p. 109-43.

[5] Loizou PC, Lobo A, Hu Y. Subspace algorithms for noise reduction in cochlear implants. *Journal of the Acoustical Society of America*. 2005;118(5):2791-3.

[6] Hu Y, Loizou PC. A subspace approach for enhancing speech corrupted by colored noise. *IEEE Signal Processing Letters*. 2002;9(7):204-6.

[7] Bolner F, Goehring T, Monaghan J, van Dijk B, Wouters J, Bleeck S, editors. *Speech enhancement based on neural networks applied to cochlear implant coding strategies*. International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2016; Shanghai, China.

[8] Scalart P, Vieira J. Speech enhancement based on a priori signal to noise estimation. *1996 IEEE International Conference on Acoustics, Speech, and Signal Processing, Conference Proceedings, Vols 1-6*. 1996:629-32.

[9] Koning R, Madhu N, Wouters J. Ideal Time-Frequency Masking Algorithms Lead to Different Speech Intelligibility and Quality in Normal-Hearing and Cochlear Implant Listeners. *IEEE Transactions on Biomedical Engineering*. 2015;62(1):331-41.

[10] Hochberg I, Boothroyd A, Weiss M, Hellman S. Effects of noise and noise suppression on speech perception by cochlear implant users. *Ear Hear*. 1992;13(4):263-71.

[11] Weiss MR. Effects of noise and noise reduction processing on the operation of the Nucleus-22 cochlear implant processor. *Journal of Rehabilitation Research and Development*. 1993;30(1):117-28.

[12] Yang LP, Fu QJ. Spectral subtraction-based speech enhancement for cochlear implant patients in background noise. *Journal of the Acoustical Society of America*. 2005;117(3):1001-4.

[13] Gustafsson H, Nordholm SE, Claesson I. Spectral subtraction using reduced delay convolution and adaptive averaging. *IEEE Transactions on Speech and Audio Processing*. 2001;9(8):799-807.

[14] Verschuur C, Lutman M, Wahat NH. Evaluation of a non-linear spectral subtraction noise suppression scheme in cochlear implant users. *Cochlear Implants International*. 2006;7(4):193-6.

[15] Lockwood P, Boudy J. Experiments with a Nonlinear Spectral Subtractor (Nss), Hidden Markov-Models and the Projection, for Robust Speech Recognition in Cars. *Speech Communication*. 1992;11(2-3):215-28.

[16] Kallel F, Frikha M, Ghorbel M, Ben Hamida A, Berger-Vachon C. Dual-channel spectral subtraction algorithms based speech enhancement dedicated to a bilateral cochlear implant. *Applied Acoustics*. 2012;73(1):12-20.

[17] Berouti M, Schwartz R, Makhoul J. Enhancement of speech corrupted by acoustic noise. *ICASSP 79 IEEE International Conference on Acoustics, Speech and Signal Processing*. 1979:208-11.

[18] Kamath S, Loizou P. A multi-band spectral subtraction method for enhancing speech corrupted by colored noise. *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing, Vols I-IV, Proceedings*. 2002:4164-7.

[19] Li JF, Yang L, Zhang JP, Yan YH, Hu Y, Akagi M, et al. Comparative intelligibility investigation of single-channel noise-reduction algorithms for Chinese, Japanese, and English. *Journal of the Acoustical Society of America*. 2011;129(5):3291-301.

[20] Dachasilaruk S, Luangsawang C, Jantharamin N, Kaewsri S, Doungta P, editors. *Evaluation of Thai speech intelligibility based on noise reduction techniques for cochlear implants*. The 10th Image and Signal Processing, BioMedical Engineering and Informatics; 2017; Shanghai, China.

[21] Kiefer J, Hohl S, Sturzebecher E, Pfennigdorff T, Gstoettner W. 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*. 2001;40(1):32-42.

[22] Skinner MW, Holden LK, Whitford LA, Plant KL, Psarros C, Holden TA. Speech recognition with the Nucleus 24 SPEAK, ACE, and CIS speech coding strategies in newly implanted adults. *Ear and Hearing*. 2002;23(3):207-23.

[23] Cochlear. ACE™ and CIS DSP Strategies. Lane Cove, New South Wales, Australia: 2002.

[24] Nogueira W, Buchner A, Lenarz T, Edler B. A psychoacoustic "NofM"-type speech coding strategy for cochlear implants. *Eurasip Journal on Applied Signal Processing*. 2005;2005(18):3044-59.

[25] Wilson BS, Dorman MF. Cochlear implants: Current designs and future possibilities. *Journal of Rehabilitation Research and Development*. 2008;45(5):695-730.

[26] Hu Y, Loizou PC. A new sound coding strategy for suppressing noise in cochlear implants. *Journal of the Acoustical Society of America*. 2008;124(1):498-509.

[27] Komalarajun S. Development of Thai Speech Discrimination Materials: Mahidol University; 1979.

[28] Kangsanarak B. Development of Thai Spondee Words for Speech Audiometry: Mahidol University; 1980.

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

[30] Chen JD, Benesty J, Huang Y, Doclo S. New insights into the noise reduction Wiener filter. *IEEE Transactions on Audio Speech and Language Processing*. 2006;14(4):1218-34.

[31] Theera-Umpon N, Chansareewittaya S, Auephanwiriyakul S. Phoneme and tonal accent recognition for Thai speech. *Expert Systems with Applications*. 2011;38(10):13254-9.

[32] Mishra C, D.L.Gupta. Deep Machine Learning and Neural Networks: An Overview. *International Journal of Artificial Intelligence (IJAI)*. 2017;6:66-73.

[33] Sulong A, Gunawan TS, Khalifa OO, Kartiwi M, Ambikairajah Eb. Speech enhancement based on Wiener filter and compressive sensing. *Indonesian Journal of Electrical Engineering and Computer Science*. 2016;2:367-79.

[34] Ning K, Qin G. Compressed sensing speech signal enhancement research. *Indonesian Journal of Electrical Engineering and Computer Science*. 2017;6(1):26-35.