Combination and Comparison of Sound Coding Strategies Using Cochlear Implant Simulation With Mandarin Speech

Enoch Hsin-Ho Huang, Chao-Min Wu, *Member, IEEE*, and Hung-Ching Lin

***Abstract*— Three cochlear implant (CI) sound coding strategies were combined in the same signal processing path and compared for speech intelligibility with vocoded Mandarin sentences. The three CI coding strategies, biologically-inspired hearing aid algorithm (BioAid), envelope enhancement (EE), and fundamental frequency modulation (F0mod), were combined with the advanced combination encoder (ACE) strategy. Hence, four singular coding strategies and four combinational coding strategies were derived. Mandarin sentences with speech-shape noise were processed using these coding strategies. Speech understanding of vocoded Mandarin sentences was evaluated using short-time objective intelligibility (STOI) and subjective sentence recognition tests with normal-hearing listeners. For signal-to-noise ratios at 5 dB or above, the EE strategy had slightly higher average scores in both STOI and listening tests compared to ACE. The addition of EE to BioAid slightly increased the mean scores for BioAid+EE, which was the combination strategy with the highest scores in both objective and subjective speech intelligibility. The benefits of BioAid, F0mod, and the four combinational coding strategies were not observed in CI simulation. The findings of this study may be useful for the future design of coding strategies and related studies with Mandarin.**

***Index Terms*— Cochlear implant, combinational sound coding strategy, vocoder, speech intelligibility, Mandarin.**

I. Introduction

THE cochlear implant (CI) is a successful neural prosthetic device that restores the sense of hearing for hundreds of thousands of hearing-impaired listeners. With an external sound processor and an internal implant, a CI bypasses a malfunctioning cochlea [1]–[3] and transmits sound signals through the electro-to-neural interface to stimulate auditory nerves [4], [5]. The coding strategy in a CI sound processor plays an important role in defining the stimulation patterns achievable and recognizable by the auditory brain. With advances in coding strategies, CI users can understand speech conversations in quiet and even use the telephone [2].

Sound coding strategies, which significantly contribute to speech comprehension in CI listening, are constantly developed by many researchers [6]. The continuous interleaved sampling (CIS) [7] and advanced combination encoder (ACE) [8]–[10] strategies using sequential electrode stimulation and N-of-M maxima selection, respectively, have successfully helped CI recipients to achieve 80-90% sentence recognition in quiet [2]. More recent strategies MP3000 [11], [12], fine structure processing (FSP) [13], and HiRes 120 or Optima [14], [15] available in commercial sound processors are based on psychoacoustic masking, pulse patterns carrying the temporal fine structure (TFS) [16]–[19], and current steering, respectively. Furthermore, several experimental coding strategies have been proposed using techniques such as preserving pitch [20]–[28], formants [29], or harmonics [30]; enhancing transient envelopes [31]–[33] or spectral contrast [34]; and applying auditory physiology [35]–[39] or binaural hearing [37]–[39] mechanisms. Today, further innovations of coding strategies are still needed to overcome the challenges in CI perception, including recognizing speech in noise and tonal languages [1]–[3], [6].

Comparative evaluations of CI coding strategies may provide new insights for innovating CI signal processing. Owing to the limited availability of source codes, only a few experimental coding strategies proposed by different research teams have been compared [6], [40], [41]. Most studies often evaluate only one or two variants of a proposed strategy with a reference strategy, such as CIS or ACE [26], [29], [34], [36]. There is a need to compare experimental strategies that were independently developed using identical experimental setups.

Combining different coding strategies may accumulate the strengths of individual approaches to produce better overall results. Because many experimental coding strategies are essentially deformations of the ACE strategy [11], [21], [29],

Manuscript received January 21, 2021; revised May 17, 2021 and August 17, 2021; accepted November 10, 2021. Date of publication November 12, 2021; date of current version November 23, 2021. This work was supported by the Ministry of Science and Technology of Taiwan under Grant MOST 107-2221-E-008-031. *(Corresponding author: Chao-Min Wu.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institution: Academia Sinica, Taipei, Taiwan; IRB Board: Institutional Review Board on Biomedical Science Research under Protocol No. AS-IRB-BM-20041, Dated September 21, 2020.

Enoch Hsin-Ho Huang and Chao-Min Wu are with the Department of Electrical Engineering, National Central University, Taoyuan 320317, Taiwan (e-mail: wucm@ee.ncu.edu.tw).

Hung-Ching Lin is with the Department of Otolaryngology, Mackay Memorial Hospital, Taipei 104217, Taiwan, and also with the Department of Audiology and Speech-Language Pathology, Mackay Medical College, New Taipei City 252005, Taiwan.

Digital Object Identifier 10.1109/TNSRE.2021.3128064

[30], [33], [34], it is possible to combine them with ACE in the same CI signal processing path.

The investigation of coding strategies with Mandarin may help understand how to improve tonal language recognition, one of the major challenges in CI listening. Mandarin, the most widely spoken tonal language and Chinese dialect, contains syllables with different intonations to represent distinct meanings. Each Mandarin syllable is coupled with one of four primary lexical tone contours, which are tone 1 (high and flat), tone 2 (rising), tone 3 (falling then rising), and tone 4 (falling), or an additional neutral tone. Mandarin speech perception has been investigated with commercial coding strategies, such as ACE [42]–[44], FSP [45], [46], and HiRes 120 [47]. Fundamental frequency (F0) processing based strategies, including frequency amplitude modulation encoding (FAME) [20], [41], fundamental frequency modulation (F0mod) [21]–[24], optimized pitch and language (OPAL) [26], and C-tone [28], have also been studied with Mandarin speech. Other experimental coding strategies without specialized designs for F0 are often evaluated only with non-tonal languages, such as English [29], Dutch/Flemish [33], German [34]–[36], and Spanish [39]. Some of these strategies may also benefit Mandarin understanding since speech recognition relates to not only lexical tones but also other phonemic information. Therefore, the performance of both F0 and non-F0 based strategies with a common Mandarin experiment is of interest for investigation.

Different coding strategies may be combined for examining their individual performance and combinational benefits with vocoded Mandarin. In this study, three approaches, biologically-inspired hearing aid algorithm (BioAid) [48]–[52], envelope enhancement (EE) [31]–[33], and F0mod [21]–[24] have been chosen because of their distinct strengths in improving electric hearing. They may be combined in the same signal processing path to form eight coding strategies, BioAid, EE, F0mod, BioAid+EE, BioAid+F0mod, EE+F0mod, BioAid+EE+F0mod, and the reference ACE strategy. To investigate the performance of both F0 and non-F0 based strategies for a tonal language without actual experiments with CI implantees, objective and subjective evaluations can be carried out using the CI simulation of Mandarin sentences. The results may also help determine whether BioAid, EE, and F0mod can accumulate their individual strengths in the derived combinational strategies. The methods and findings of this study may provide some new insights into CI signal processing and Mandarin speech recognition.

This article is organized as follows. Section II describes the implementation, combination, and evaluation of the coding strategies. Section III compares the results of objective prediction and listening tests. Section IV discusses the experimental results. Section V describes the conclusions of the investigation.

II. Methods

This section describes the CI signal processing, simulation platform, speech materials, singular and combinational coding strategies, electrode stimulation patterns of the strategies, and methods for objective and subjective evaluation.

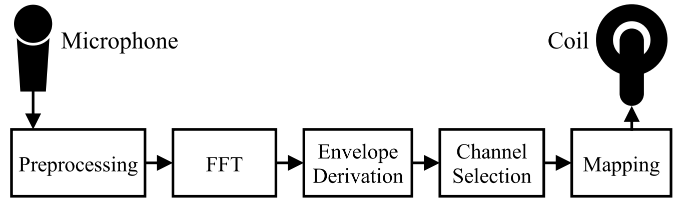

Fig. 1. Signal processing stages of a CI sound processor.

A. CI Signal Processing and Simulation

A CI sound processor involves a series of signal processing stages as depicted in Fig. 1. Acoustic waves are collected by microphones and converted into sound signals. The preprocessing stage enhances sound features or reduces interference using signal processing techniques such as microphone beamformer, pre-emphasis, automatic gain control (AGC), scene classifiers, and speech enhancement algorithms [53]–[55]. A filter bank or a Fast Fourier Transform (FFT) step converts preprocessed sounds into multi-band signals, which are further grouped together to derive the envelopes for different electrodes [11]. A maxima channel selection or N-of-M stage outputs the N channels with the largest amplitudes from the M channels for every sound frame during processing. Spectral contents of the M selected channels are adjusted in magnitude by a loudness growth function (LGF) and then mapped to current levels according to a CI recipient's threshold and comfortable level (T-level and C-level) settings. Each channel typically corresponds to a physical electrode. The resulting electrode signals are converted into radio frequencies (RF) for transmitting to the internal implant for stimulating the auditory nerves of the CI recipient. To estimate the performance of electric hearing without CI recipients, the user-dependent mapping stage and the RF coil in Fig. 1 can be replaced by a vocoder [56]. By modulating tone or noise carriers using the channel envelopes via a vocoder, stimulations pulses are converted to audio signals for listening with earphones or loudspeakers.

B. NCU-CI and Speech Materials

NCU-CI, a software-based CI simulation platform previously developed by the National Central University (NCU) [57], [58], has been used for coding strategy implementation and subjective listening tests in normal-hearing (NH) listeners. Based on MATLAB version R2020a, NCU-CI provides options for the experimenter to choose speech materials, bimodal (CI + HA) or bilateral CI modes, noise type, signal-to-noise ratio (SNR), and CI parameters including the coding strategy, vocoder carrier, stimulation rate, and the number of maxima channels. ACE, CIS, and several other experimental coding strategies are included in the platform. A first-order Butterworth highpass filter with a cutoff frequency of 1,200 Hz is used as a pre-emphasis filter. The FFT converts samples multiplied by a 128-point Hann window into 65 bands, which are weighted and grouped into envelopes of

22 channels following the descriptions in [11]. Typical ACE settings of 8 maxima channels were used for all simulations in this study.

NCU-CI has a graphical user interface in traditional Chinese and Mandarin speech materials including monosyllabic [59], bisyllabic [60], and trisyllabic [61] words, as well as the Taiwan Mandarin Hearing in Noise Test (TMHINT) sentences [62]. The TMHINT speech material involves 320 sentences (16 lists $\times$ 20 sentences/list) and each sentence consists of 10 syllables. Each list is balanced in phonemes [62]. The voice recordings of TMHINT are speech of one male speaker and one female speaker downsampled to 16 kHz and normalized to 65 dB SPL. Speech-shaped noise (SSN) can be added to speech at -10, -5, 0, 5, 10, and 15 dB SNRs. CI simulation sounds are synthesized using a vocoder [56]. In the practice mode, sentences of a fixed list are presented. In the test mode, sentences of a randomly selected list are presented, and the user interface may be used to type in recognized Chinese characters for further calculations of the speech intelligibility.

C. BioAid, EE, and F0mod

In the present study, three sound coding strategies, BioAid, EE and F0mod, were implemented in NCU-CI. These strategies are described as below.

BioAid is a biologically inspired hearing aid algorithm based on two auditory mechanisms, dynamic compression and auditory efferent suppression, to improve frequency selectivity for hearing-impaired listeners [48]–[52]. The first auditory mechanism, the instantaneous compressive property of the basilar membrane (BM), is mimicked by the dual resonance nonlinear (DRNL) filterbank [63] with broken-stick compression and bandpass filtering. The DRNL filterbank consists of a linear pathway and a nonlinear pathway representing the mechanical and active processes of the cochlea, which are physiologically controlled by the inner hair cells and outer hair cells along the BM, respectively [64]. The second mechanism of the ipsilateral medial olivocochlear (MOC) reflex, the auditory efferent suppression induced by the brainstem reflex, is modeled as a feedback loop called delayed feedback attenuation control (DFAC). The DFAC works as a delayed automatic gain control (AGC) to adjust the sound levels entering the DRNL filterbank based on feedback signals and control parameters such as feedback delay time, integration time constant, signal threshold, and compression ratio. The DRNL signals are processed to form audio signals through a series of bandpass filters, within-channel gains, and a summation of different channels. Langner and Jürgens [50] demonstrated improved forward-masked frequency selectivity with BioAid in both CI simulations with NH listeners and CI recipients by measuring the psychophysical tuning curve (PTC). While BioAid is a hearing aid compression algorithm, it is referred to as a coding strategy in this study.

Envelope enhancement (EE) is a coding strategy that enhances transient speech variations to improve speech understanding in noise [31]–[33]. Compared to the ACE strategy, EE uses an additional peak signal derivation module in each channel to enhance fast transitions of speech contents, such as transients and onsets, which are typically with high entropy [32]. However, these important contents are often not represented well in slow-varying CI envelopes using electrical stimulation lacking the adaptation effect at synapses between inner hair cells (IHCs) and auditory nerve fibers (ANFs) [31]. Therefore, EE attempts to mimic and compensate for the adaptation effect across the CI channels. The peak signals related to rapid speech variations are extracted using peak signal derivation modules consisting of lowpass filters, half-rectifiers, and amplifiers. Extracted peaks are added back to the original channel envelopes before channel selection. Koning and Wouters (2016) showed that the EE strategy can improve the recognition of keywords in Dutch sentences below 6 dB SNR and the stop consonant reception at 6 dB SNR and in quiet compared to ACE.

Fundamental frequency modulation (F0mod) is a strategy designed to enhance pitch information across CI channels, and hence improve music and tonal language recognition [21]–[24]. For electric hearing, the pitch can be perceived using place cues and temporal cues [22]. In coding strategies with fixed pulse rates such as ACE, place pitch cues are represented by stimulating the tonotopical sites along the multi-electrode array, while temporal pitch cues are embedded in pulse trains transmitted by each electrode. F0mod aims to enhance the periodicity of the temporal pitch cues in the slow-varying envelopes over different electrodes. A fundamental frequency (F0) detector, based on the autocorrelation function (ACF), is used to extract the pitch frequency of speech signals. Sound segments are classified as unvoiced if the detected pitch falls outside the valid range of speech F0, and processed identically to the ACE strategy. Voiced frames, however, are manipulated in pulse magnitudes so that the slow-varying channel envelopes are amplitude modulated at the detected F0. In the listening tests with Mandarin-speaking CI recipients conducted by Milczynski *et al.* [23], better lexical tone perception was observed with F0mod than with ACE, but the sentence recognition results were similar.

D. Combination and Implementation

The three coding strategies, BioAid, EE, and F0mod, were combined and implemented in the same signal processing path. As illustrated in the dotted line blocks in Fig. 2, BioAid, EE, and F0mod were treated as independent sound processing stages, excluding the preprocessing and channel selection stages. Similar to the design by Langner and Jürgens [50], BioAid was considered as an independent preprocessing stage in front of the FFT in this study. The EE stage was located behind the envelope detection stage to enhance transient variations of envelopes before further modulation by the F0mod stage at F0's detected from speech signals before FFT. In terms of auditory physiology, EE simulates the fast adaptation of synapses between inner hair cells and auditory nerves [31], and hence the enhanced signals can be can further processed by F0mod to enrich temporal pitch information for stimulation [66]. The BioAid algorithm was built using the source code available online [65], while the EE and F0mod strategies were implemented within the present study

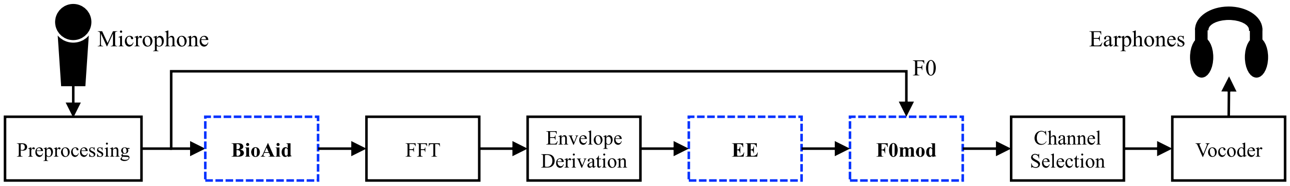

Fig. 2. Combination of the BioAid, EE, and F0mod processing stages to form new coding strategies for CI simulation. Independently activating or deactivating the BioAid, EE, and F0mod stages in the blue dotted line blocks may derive eight strategies, including four *singular strategies* (ACE, BioAid, EE, and F0mod) and four *combinational strategies* (BioAid+EE, BioAid+F0mod, EE+F0mod, and BioAid+EE+F0mod). ACE is the condition that all the three blue dotted line blocks are bypassed. BioAid, EE, and F0mod are considered as singular strategies including the preprocessing and channel selection stages. The vocoder and earphones are used for simulation with NH listeners. F0mod detects the fundamental frequency (F0) from the preprocessed signals.

according to [21]–[24], [31]–[33]. A noise vocoder was used to generate acoustic CI simulation for presenting to NH listeners via earphones as shown in Fig. 2, instead of producing electric stimulations using a mapping function and a radio frequency coil as illustrated in Fig. 1.

The processing stages of the BioAid, EE, and F0mod could be independently activated or deactivated to form a total of eight strategies as in Fig. 2. When all three approaches were bypassed, the signal processing path was the ACE strategy. By enabling only one of the three blue dotted line blocks of processing stages in Fig. 2, the CI strategy became BioAid, EE, or F0mod. In this study, the three strategies and ACE were referred to as the four *singular coding strategies* including the preprocessing and the N-of-M channel selection stages. In other words, the combinations of the BioAid, EE, or F0mod processing stages with channel selection were defined as the BioAid, EE, or F0mod strategy, but not BioAid+ACE, EE+ACE, or F0mod+ACE. By activating two or three processing stages of BioAid, EE, and F0mod, four *combinational coding strategies* of BioAid+EE, BioAid+F0mod, EE+F0mod and BioAid+EE+F0mod were obtained. Consequently, eight coding strategies can be derived with the implementation shown in Fig. 2.

The three singular strategies, BioAid, EE, and F0mod, were customized for NCU-CI. The C++ version of BioAid was incorporated into MATLAB codes via the MEX API interface [65]. The within-channel gain at the last stage before channel summation was deactivated as in [50]. A fixed instantaneous compression threshold was applied to all BioAid bands, while the other parameters generally followed the default settings in [65]. The EE strategy was developed to fit the acoustic properties of the TMHINT sentences. Koning and Wouters (2016) used both clean speech and noisy speech with an interfering speaker in peak detection, but the proposed approach detected peaks from noisy speech only. In contrast to the previous F0mod study that detected F0 from clean speech [23], a more realistic approach was adopted to determine pitch directly from noisy speech using an ACF-based F0 detector. Sound segments with pitch frequencies detected between 50 and 500 Hz were labeled as voiced frames for amplitude modulating the channel envelopes at the detected F0.

E. Electrode Stimulation Patterns

To understand the electrode stimulation patterns of the implemented coding strategies, the electrodograms and the reference spectrogram are shown in Fig. 3. The bisyllabic Mandarin phrase “Xuǎn Zé” (“Choice” in Chinese with tones 3 and 2) consists of two syllables separated by a silent segment at about 0.5s. The spectrogram of the unprocessed bisyllable depicted in Fig. 3(a) is a reference for details of the original speech. The electrodograms of five coding strategies are illustrated using the sequence plotting function in the Nucleus Matlab Toolbox (NMT) [67]. In Fig. 3(c), the initial vowel envelopes indicated by orange rectangles are preserved with BioAid compared to those with ACE in Fig. 3(b), while subsequent pulses are suppressed in amplitudes probably because of the attenuation mechanism controlled by the DFAC feedback loop. As for EE in Fig. 3(d), the pulses for both syllables are generally similar to the ACE stimuli, and the amplitudes are more pronounced at some high frequency electrodes indicated by the blue rectangles. The additional stimulation of the EE strategy in the blue rectangles do not occur in the electrodogram with the ACE strategy. In Fig. 3(e), the sparse pulses for F0mod are with less density compared to the saturated ACE stimuli. By zooming in on the second syllable “Zé” in Fig. 3(b) for ACE and in Fig. 3(e) for F0mod, the resulting electrodograms in Fig. 4 with electrode frequencies in the range of channel 2 (375 Hz) to channel 6 (875 Hz) show the different pulse patterns generated by the two strategies. The F0mod pulses with periodic rise and fall indicate the effect of amplitude modulation. The color rectangles in Fig. 3(f) show that BioAid+EE preserves some characteristics of singular strategies BioAid and EE with the combination of similar pulse patterns in Fig. 3(c) and Fig. 3(d).

F. Objective and Subjective Evaluation

To estimate the speech intelligibility in CI users, objective and subjective evaluations using vocoded sentences were designed. The results were analyzed using IBM SPSS Statistics 21 (released 2012) with the significance level $\alpha = 0.05$.

Speech materials were prepared for evaluations. TMHINT sentences pronounced by a female speaker and by a male speaker were added with SSN at seven SNR levels (−10, −5, 0, 5, 10, 15 dB SNRs and quiet). The noisy sentences were processed by the eight coding strategies to form electrode stimulation patterns, which were transformed into simulated speech using a noise vocoder [56] and then normalized in sound pressure level.

To compare the performance of the eight coding strategies, speech intelligibility was estimated using an objective

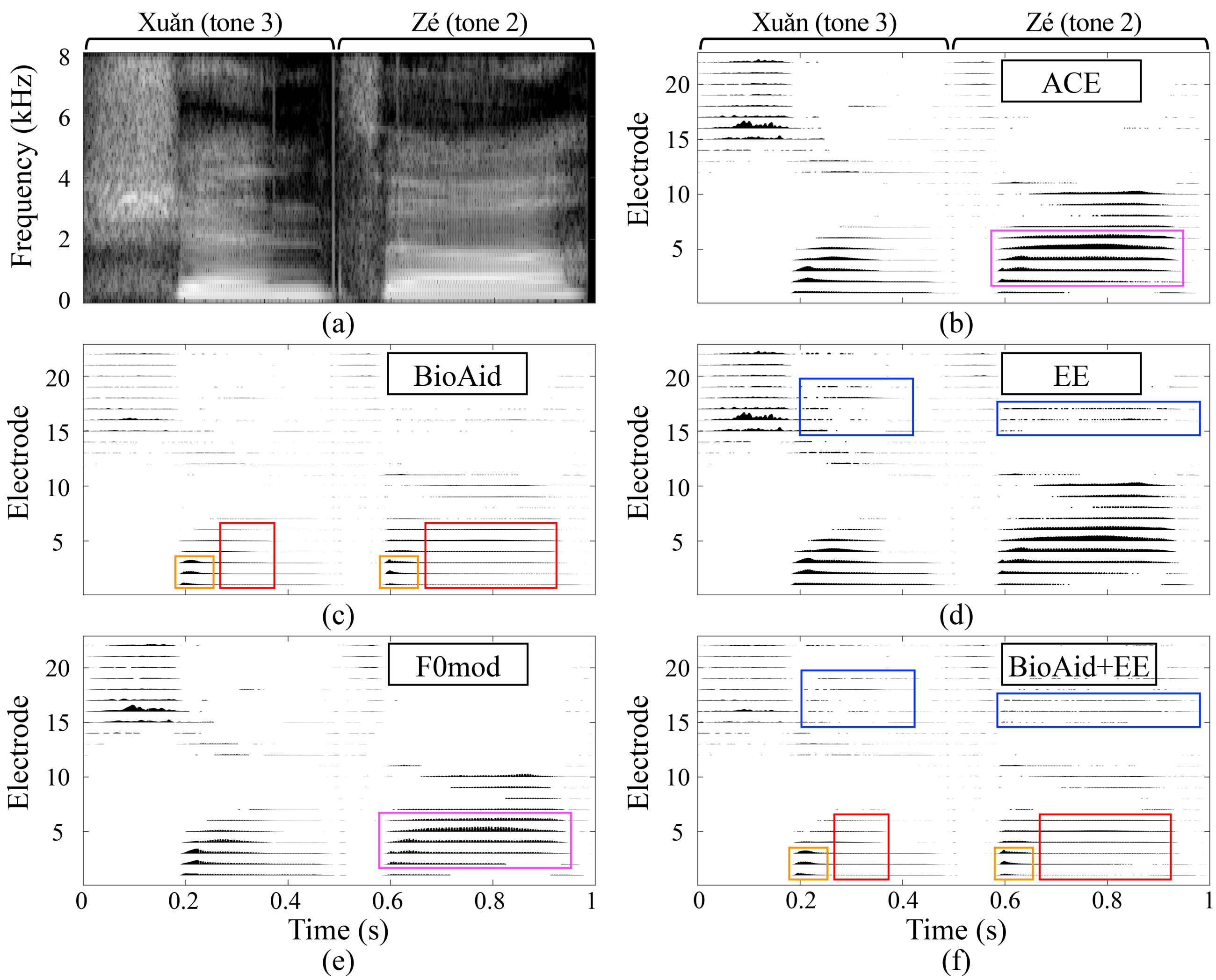

Fig. 3. (a) Spectrogram for the unprocessed Mandarin bisyllabic phrase "Xuǎn Zé" ("Choice" in Chinese with tones 3 and 2). Electrodograms for the speech processed by coding strategies (b) ACE, (c) BioAid, (d) EE, (e) F0mod, and (f) BioAid+EE. BioAid preserves the initial vowel envelopes in the orange rectangles as ACE, but suppresses the subsequent pulses in red rectangles. EE provides additional stimulation not present or prominent in the representation with ACE as indicated with blue rectangles. The pink rectangles showing more sparse pulses with F0mod than with ACE are further enlarged in Fig. 4. BioAid+EE produces pulse patterns similar to BioAid in orange and red rectangles and similar to EE in blue rectangles.

prediction measure. Several objective intelligibility prediction measures which have been validated using hearing aids and CIs [68]–[70]. Short-time objective intelligibility (STOI) [71] is a popular objective intelligibility prediction metric, which have been used in several CI studies [38], [72]–[77]. By computing the relation between clean and degraded speech, the STOI score obtained in the range of 0 to 1 has a positive correlation with the speech intelligibility in noise [71]. In this study, STOI scores were obtained between unprocessed and vocoded sentences.

While STOI scores were measured for all sentences under every SNR and strategy condition, the amounts of speech materials in the subjective evaluation were reduced to simplify the listening test procedure. Only female sentences at 0 dB and 5 dB SNRs were used because the previous experiences with NCU-CI showed that the sentence recognition scores across various SNRs were considerably changed at these two SNRs [57]. Furthermore, only five coding strategies were tested, including the four singular strategies (ACE, BioAid, EE, and F0mod), and BioAid+EE, the combinational strategy

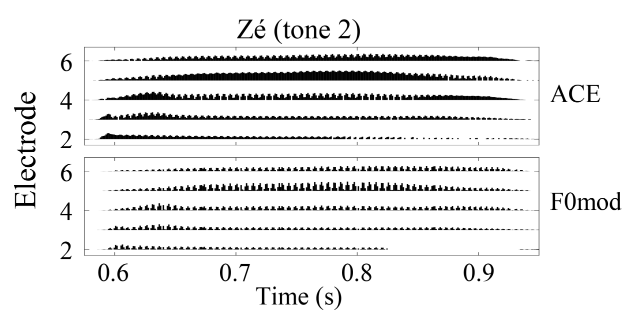

Fig. 4. Electrodogram for the "Zé" syllable (enlarged from the pink rectangles in Fig. 3) processed by ACE and F0mod between electrode 2 (375 Hz) and electrode 6 (875 Hz). The amplitudes of F0mod pulses periodically change at the detected F0.

with the highest STOI scores observed in the objective evaluation.

Subjective listening tests were carried out with NH and native Mandarin-speaking listeners under 10 conditions (5 strategies × 2 SNR levels). Vocoded sentences were presented to each subject at 65 dB SPL via Telephonics

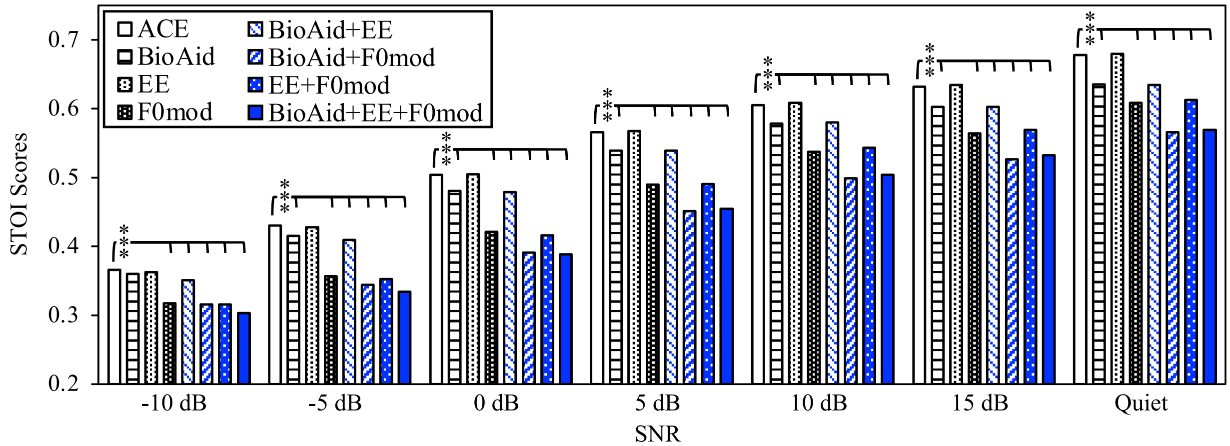

Fig. 5. Average STOI scores for 320 female TMHINT sentences at seven SNR levels processed by eight coding strategies using a CI vocoder. The results of the four combinational coding strategies are illustrated in blue. The STOI scores for the reference ACE strategy at various SNR levels are 0.366 (−10 dB), 0.431 (−5 dB), 0.504 (0 dB), 0.566 (5 dB), 0.605 (10 dB), 0.632 (15 dB) and 0.678 (quiet). The three asterisks indicates significant differences ($p < 0.001$) between ACE and various coding strategies.

TDH-39P earphones in a sound isolation booth. The hearing levels of seven male subjects aged 22-26 years were no more than 20 dB HL between 125 Hz and 8 kHz with pure tone audiometry (PTA) conducted by an audiologist. The subjects were instructed to use NCU-CI for the experiment and ask questions when needed. Each subject was familiarized with CI vocoded speech with the defined sound level in the practice session and was encouraged to type words to the best of their ability in accordance with their perception. The lists and sentences of the TMHINT material and the five coding strategies were proceeded in random order during the tests. In each test condition, sentences of a list were presented to the left ear of each subject and then a different list was presented to the right ear. After testing for the five strategies at 0 dB SNR, a short break was given before starting the 5 dB SNR session. The sentence recognition test took about 90-120 minutes, and it was approved in advance by a local institutional review board (IRB).

After the listening test, the percent correct recognition of Mandarin sentences was calculated. As previously described, TMHINT sentences have an identical length of 10 syllables. For each test condition of a strategy and an SNR level, speech intelligibility was calculated as the average percentage of correct result for both ears, and the result for each ear was computed from 200 syllables (10 syllables/sentence × 20 sentences/list). Scores were given for each correct syllable, including a homophone, a different Chinese character with the correct pronunciation.

III. Results

The objective STOI scores and the subjective listening test results are presented in this section.

A. Objective Intelligibility Prediction

The average STOI scores of 320 female TMHINT sentences under various test conditions are shown in Fig. 5. A two-way analysis of variance (ANOVA) was conducted with coding strategy and SNR level as factors. To simplify the statistical analysis, the quiet condition was treated as one of the seven SNR levels instead of being considered as the third factor of either quiet or noisy condition. For the 17,920 sentences processed (320 sentences × 8 strategies × 7 SNR levels), significant effects were observed with both coding strategy ($F[7, 17864] = 3133.8$, $p < 0.001$) and SNR level ($F[6, 17864] = 24113.8$, $p < 0.001$). The Tukey post-hoc test showed that BioAid, F0mod, BioAid+EE, BioAid+F0mod, EE+F0mod, and BioAid+EE+F0mod were significantly different to ACE ($p < 0.001$). Comparable performance was observed between ACE and EE, between BioAid and BioAid+EE, between F0mod and EE+F0mod, and between BioAid+F0mod and BioAid+EE+F0mod ($p > 0.05$). Post-hoc analysis at each SNR level showed significant differences for the majority of strategies in comparison to ACE ($p < 0.001$) in Fig. 5, except for EE at all SNR levels and for BioAid at −10 dB SNR ($p > 0.05$). Therefore, the addition of the EE processing stage to ACE, BioAid, F0mod, and BioAid+F0mod did not affect the statistical significance of the STOI scores. For male sentences, the results not shown were in similar trends to those in Fig. 5.

In Fig. 5, ACE and EE are the two coding strategies with the highest STOI scores at all SNR levels for female sentences. The mean results for EE were slightly greater than those for ACE within a range of 0.001-0.003 at 0 dB SNR or above, but smaller by 0.003 at both −5 dB and −10 dB SNRs. For male sentences, the mean STOI scores were also very close between EE and ACE. No significant differences were found between EE and ACE for both female and male speech ($p > 0.05$).

The STOI scores for BioAid and F0mod were not as high as those for ACE. Across the seven SNR levels, the mean STOI scores of female sentences processed by BioAid and F0mod were less than those for ACE in the range of 0.005-0.042 and 0.048-0.083, respectively. The results for male speech showed a similar trend between these strategies.

Combinational coding strategies generally resulted in lower STOI scores than singular coding strategies did, except for the strategies including EE. The four combinational strategies, BioAid+EE, BioAid+F0mod, EE+F0mod, and BioAid+EE+F0mod, were unable to achieve results close

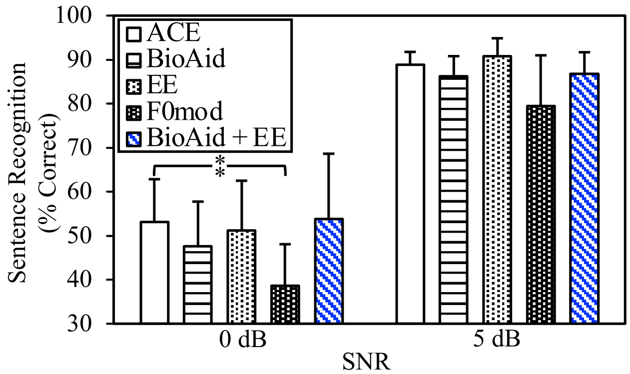

Fig. 6. Average Mandarin sentence recognition scores and standard deviations of seven NH subjects for ACE, BioAid, EE, F0mod and BioAid+EE at two SNR levels. The percent correct scores for the reference ACE strategy are 53.1% at 0 dB and 88.9% at 5 dB SNR. The results of the combinational strategy BioAid+EE are illustrated in blue. The highest mean scores at 0 dB and 5 dB SNRs are 53.8% for BioAid+EE and 90.7% for EE, respectively. The two asterisks denote a significant difference ($p < 0.01$) between ACE and F0mod at 0 dB SNR.

to those of ACE. However, the combinations of the EE processing stage did not degrade the performance. This was observed from the comparable results between BioAid and BioAid+EE, between F0mod and EE+F0mod, and between BioAid+F0mod and BioAid+EE+F0mod ($p > 0.05$). At 5 dB SNR or above, the mean STOI scores were slightly greater for BioAid+EE than BioAid, for EE+F0mod than F0mod, and for BioAid+EE+F0mod than BioAid+F0mod. Similar outcomes were obtained with male sentences. By comparing the objective intelligibility results with both female and male sentences, all of the four combinational strategies were significantly different to ACE ($p < 0.001$), while the combination of the EE stage did not change the results significantly ($p > 0.05$).

To understand the relative performance, the eight coding strategies can be listed in decreasing order of mean STOI scores. For female sentences at 5dB SNR or above, the descending order was EE, ACE, BioAid+EE, BioAid, EE+F0mod, F0mod, BioAid+EE+F0mod, and BioAid+F0mod. At 0 dB SNR, the order became EE, ACE, BioAid, BioAid+EE, F0mod, EE+F0mod, BioAid+EE, and BioAid+EE+F0mod. As the SNR decreased to −5 dB or below, ACE resulted in the highest scores and outperformed EE, while the order of subsequent strategies remained unchanged. For male sentences, the relative orders of STOI scores were generally similar to those for female sentences. BioAid+EE was the combinational strategy with the highest STOI scores at all SNR levels, and hence was to be used for the subjective listening test.

B. Subjective Listening Test

The subjective listening test results for female speech materials processed by the five coding strategies are shown in Fig. 6. A two-way repeated-measures (RM) ANOVA was conducted. Significant effects were observed with both factors of coding strategy ($F[4, 24] = 7.189$, $p = 0.001$) and SNR ($F[1, 6] = 155.908$, $p < 0.001$). The precent correct results for the five strategies at 5 dB SNR were all higher than those at 0 dB SNR. Post-hoc pairwise comparisons with Bonferroni adjustments at each SNR showed that ACE and F0mod were significantly different at 0 dB SNR ($p < 0.01$).

Apart from ACE, EE was the singular coding strategy with the highest recognition results. The results between ACE and EE at both 0 and 5 dB SNRs were not significantly different ($p > 0.05$). The scores of the reference ACE strategy were 53.1% at 0 dB SNR and 88.9% at 5 dB SNR. The average percent correct score for EE at 0 dB SNR was 51.1% and only 2% less than the score for ACE. At 5 dB SNR, the result for EE was 90.7%, which was the highest outcome among the five strategies and 1.8% higher than that of ACE.

The sentence recognition scores for BioAid were statistically similar to ACE at both SNR levels ($p > 0.05$). The average results for BioAid were 47.5% at 0 dB and 86.2% at 5 dB SNR, and slightly lower than those for ACE by 2.7% and 5.6%, respectively.

The result for F0mod was significantly different to the result for ACE at 0 dB SNR ($p < 0.01$), but the two strategies were comparable at 5 dB SNR ($p > 0.05$). The mean percentage of correct scores for F0mod were 38.6% at 0 dB and 79.5% at 5 dB SNR. The percentage differences between ACE and F0mod were 14.5% at 0 dB and 9.4% at 5 dB SNR.

For the combinational coding strategy BioAid+EE, the precent correct scores at both 0 and 5 dB SNRs were not significantly different to those for ACE, BioAid and EE ($p > 0.05$). At 0dB SNR, the 53.8% recognition of BioAid+EE was the top score among the results of the five strategies and slightly higher than the result of ACE by 0.7%. As for 5 dB SNR, the recognition score for BioAid+EE was 86.8%, about 2.1% less than the score of ACE. BioAid+EE increased the mean results of BioAid by 0.5% at 0 dB SNR and 6.2% at 5 dB SNR.

The subjective results for the five coding strategies were summarized as below. The descending order of the strategies in terms of percent correct results was BioAid+EE, ACE, EE, BioAid, and F0mod at 0 dB SNR, and EE, ACE, BioAid+EE, BioAid, and F0mod at 5 dB SNR. The average percent correct scores for ACE, BioAid, EE, and BioAid+EE were within ranges of 47-54% at 0 dB SNR and 86-91% at 5 dB SNR. Apart from F0mod, the differences between ACE and all the other three strategies were less than 6% at both 0 and 5 dB SNRs and statistically insignificant ($p > 0.05$). At 0 dB SNR, the mean percentage of correct scores for ACE, EE, and BioAid+EE were all above 50%.

IV. Discussion

This section examines the results for singular and combinational coding strategies and provides some general discussions.

A. Singular Coding Strategies

BioAid was statistically different to ACE in STOI scores above −10 dB SNR ($p < 0.05$), but comparable in listening tests ($p > 0.05$). The average results for BioAid were lower than those for ACE within ranges of 0.005-0.042 in STOI scores and 2.6-5.5% in sentence recognition scores. The degraded speech intelligibility could be due to the nonlinear compression of BioAid, which caused speech distortions. Previous studies have evaluated BioAid for its frequency selectivity [49], [50] and compressive properties [51]. Only one study has investigated the speech intelligibility of hearing aid users by measuring the speech reception threshold (SRT) with binaurally linked BioAid-based dynamic compression (BADC) [52]. Based on the contributions of improved frequency selectivity for CI users in [50], further investigation may improve BioAid with CIs in reducing distortions while maintaining or even improving speech intelligibility.

A relative good performance was observed for EE with vocoded Mandarin speech. In a previous study [33], significant improvements were observed with the EE strategy at −2, 2 and 6 dB SNRs by measuring correct keywords in Dutch sentences with CI subjects. The capability of EE was also demonstrated for the Dutch words “de stad” (“the city”) [33]. Interestingly, even for Mandarin speech containing very different onsets, syllables, and tone contours to those in Dutch, the present study observed slightly higher average STOI scores for EE than ACE with female Mandarin sentences at 0 dB SNR or above. At 5 dB SNR, the subjective evaluation showed slightly higher results with EE (90.7%) than with ACE (88.9%), and six out of the seven NH subjects performed better with EE. Although the differences were small, a consistent trend was shown in both objective and subjective evaluations. The increased intelligibility of EE may be due to enhanced spectral contrast between amplitude peaks and valleys [34], [78], [79], thus some important phonemic cues such as fricatives or rapid tone changes may have been emphasized. Therefore, coding strategies not specifically designed for tonal languages may be worthy of investigation for their unforeseen advantages in Mandarin recognition.

In the present CI simulation, F0mod resulted in a slightly different performance in comparison to previous findings in CI subjects. Milczynski *et al.* [23] observed no significant differences between F0mod and ACE in recognizing male sentences. In this study, F0mod performed similarly to ACE in sentence recognition results at 5 dB SNR ($p > 0.05$), but not at 0 dB SNR ($p < 0.01$) and in STOI scores at all SNRs ($p < 0.001$). The different outcomes here could be due to the use of a pitch detector in noise and a CI vocoder. The F0 extracted from noisy speech may not be as reliable as the use of clean speech in [23], and a pitch detector with a robust feature extraction technique under interference can be considered [25], [26], [41]. Furthermore, the vocoder processing may not properly provide the periodic cues via pulse amplitudes that are known to be beneficial for CI listeners [80], [81]. Hence, further innovations for objective predictors are still required for F0-based strategies that may help improve tonal language intelligibility.

B. Combinational Coding Strategies

The speech intelligibility of combinational strategies was generally not as high as that of ACE and the other singular strategies, but some surprising exceptions have been discovered with EE. In the objective intelligibility prediction, the STOI scores of BioAid, F0mod, or BioAid+F0mod at 5, 10, and 15 dB SNRs were slightly increased after combining EE. These increases in STOI scores were similar to those obtained by comparing the EE and ACE strategies in less noisy conditions. At 5 dB SNR, similar results in sentence recognition tests showed higher recognition scores for EE than for ACE by 1.9% and for BioAid+EE than for BioAid by 0.5%. At 0 dB SNR, the scores for BioAid+EE were higher compared to ACE by 0.7%, to BioAid by 6.2%, to EE by 2.6%, and to F0mod by 15.2%. The relatively higher scores for BioAid+EE may encourage further investigations into combining EE with other well-performing strategies. The results from this study developed a framework for testing various CI coding strategies in tandem that can be expanded to other combinations of strategies, not just EE.

C. General Discussions

The improvement of Mandarin sentence recognition using CIs is challenging for many reasons. For Mandarin, speech information needs to be identified with correct combinations for all of lexical tones, consonants, and vowels. In terms of signal processing, listening in noisy environments may require innovative speech enhancement methods and sound coding strategies. While the progress in tonal language perception with coding strategies is generally limited in recent years, alternative directions for CI innovation can be considered. For example, if coding strategies can be appropriately combined, their small individual strengths may accumulate and lead to a more significant breakthrough. The performance of non-F0 based experimental coding strategies such as EE may also be further investigated with Mandarin speech.

CI simulation has both strengths and limitations. STOI may help quickly understand the relative performance between different strategies, but it may also be over sensitive to the minor distortion caused by the nonlinear process of BioAid and F0mod. The significant differences between the majority of strategies in Fig. 5 may indicate that statistical analysis may not be robust enough for all comparisons, because small differences may become significant in calculating a large number (N = 320 sentences) of STOI scores. However, comparable performance in STOI scores was still discovered between ACE and EE ($p > 0.05$). Furthermore, STOI scores and CI simulation results of NH subjects may not be accurate enough to represent the actual performance of CI users. F0mod and other signal processing approaches with remarkable changes to pulse distributions may require better objective measures and subjective listening test approaches for estimating speech intelligibility of CI listeners.

To understand the actual performance of different coding strategies with Mandarin, clinical evaluation with CI recipients will be conducted in the future. Some important experimental factors, such as stimulation intensity between strategies and parameters for individual subjects, should be carefully considered.

V. Conclusion

This study proposes a paradigm in the implementation, combination, and evaluation of CI sound coding strategies. The BioAid, EE, and F0mod strategies were arranged as intermediate stages in the same CI signal processing path to derive eight distinct strategies. The individual and combinational performance of the strategies were assessed using STOI scores and sentence recognition tests. For the first time, BioAid and EE were investigated with Mandarin speech. All the eight strategies were evaluated using CI simulation.

Coding strategies EE and BioAid+EE had slightly higher mean scores in objective and subjective speech intelligibility under less noisy conditions ($\geq$ 5 dB SNR) compared to ACE and BioAid. Similar results were observed for STOI scores when EE was combined with F0mod or BioAid+F0mod. Although not developed for tonal languages, EE and some non-F0 based CI coding strategies can still be beneficial to Mandarin speech. Furthermore, the addition of EE as a processing stage to other coding strategies may improve the overall speech intelligibility.

This study provides some valuable findings. Although the CI simulation results did not reveal the strengths of BioAid and F0mod observed in previous research, a CI vocoder may still be used with some strategies such as ACE and EE. The combination of sound coding strategies did not clearly show the accumulated benefits of the different strategies used in this study. However, the methods of combining coding strategies and speech intelligibility evaluation may still provide some insights into related fields of CI signal processing.

References

[1] B. S. Wilson and M. F. Dorman, "Cochlear implants: Current designs and future possibilities," *J. Rehabil. Res. Dev.*, vol. 45, no. 5, pp. 695–730, Dec. 2008.

[2] F.-G. Zeng, S. Rebscher, W. V. Harrison, X. Sun, and H. Feng, "Cochlear implants: System design, integration, and evaluation," *IEEE Rev. Biomed. Eng.*, vol. 1, pp. 115–142, 2008.

[3] G. M. Clark, "The multi-channel cochlear implant: Multi-disciplinary development of electrical stimulation of the cochlea and the resulting clinical benefit," *Hearing Res.*, vol. 322, pp. 4–13, Apr. 2015.

[4] G. M. Clark, "Personal reflections on the multichannel cochlear implant and a view of the future," *J. Rehabil. Res. Develop.*, vol. 45, no. 5, pp. 651–694, Dec. 2008.

[5] F.-G. Zeng, "Challenges in improving cochlear implant performance and accessibility," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 8, pp. 1662–1664, Aug. 2017.

[6] J. Wouters, H. J. McDermott, and T. Francart, "Sound coding in cochlear implants: From electric pulses to hearing," *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 67–80, Mar. 2015.

[7] B. S. Wilson, C. C. Finley, D. T. Lawson, R. D. Wolford, D. K. Eddington, and W. M. Rabinowitz, "Better speech recognition with cochlear implants," *Nature*, vol. 352, pp. 236–238, Jul. 1991.

[8] A. E. Vandali, L. A. Whitford, K. L. Plant, and A. G. M. Clark, "Speech perception as a function of electrical stimulation rate: Using the nucleus 24 cochlear implant system," *Ear Hearing*, vol. 21, no. 6, pp. 608–624, Dec. 2000.

[9] J. Kiefer, S. Hohl, E. Stürzebecher, T. Pfennigdorff, and W. Gstöettner, "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*, vol. 40, no. 1, pp. 32–42, 2001.

[10] B. L. Fetterman and E. H. Domico, "Speech recognition in background noise of cochlear implant patients," *Otolaryngol.–Head Neck Surg.*, vol. 126, no. 3, pp. 257–263, Mar. 2002.

[11] W. Nogueira, A. Büchner, T. Lenarz, and B. Edler, "A psychoacoustic 'NofM'-type speech coding strategy for cochlear implants," *EURASIP J. Adv. Signal Process.*, vol. 2005, no. 18, pp. 3044–3095, 2005.

[12] A. Bächner, W. Nogueira, B. Edler, R.-D. Battmer, and T. Lenarz, "Results from a psychoacoustic model-based strategy for the nucleus-24 and freedom cochlear implants," *Otol. Neurotol.*, vol. 29, no. 2, pp. 189–192, 2008.

[13] D. Riss *et al.*, "Envelope versus fine structure speech coding strategy: A crossover study," *Otol. Neurotol.*, vol. 32, no. 7, pp. 1094–1101, 2011.

[14] M. Brendel, A. Buechner, B. Krueger, C. Frohne-Buechner, and T. Lenarz, "Evaluation of the harmony soundprocessor in combination with the speech coding strategy HiRes 120," *Otol. Neurotol.*, vol. 29, no. 2, pp. 199–202, 2008.

[15] M. A. de Jong, J. J. Briaire, and J. H. Frijns, "Take-home trial comparing fast Fourier transformation-based and filter bank-based cochlear implant speech coding strategies," *Biomed. Res. Int.*, vol. 30, no. 2, pp. 146–152, 2009.

[16] O. Strelcyk and T. Dau, "Relations between frequency selectivity, temporal fine-structure processing, and speech reception in impaired hearing," *J. Acoust. Soc. Amer.*, vol. 125, no. 5, pp. 3328–3345, 2009.

[17] B. C. J. Moore, "The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people," *J. Assoc. Res. Otolaryngol.*, vol. 9, no. 4, pp. 399–406, Dec. 2008.

[18] B. C. Moore, *Auditory Processing of Temporal Fine Structure: Effects of Age and Hearing Loss*. Singapore: World Scientific, 2014.

[19] B. C. J. Moore, "The roles of temporal envelope and fine structure information in auditory perception," *Acoust. Sci. Technol.*, vol. 40, no. 2, pp. 61–83, 2019.

[20] K. Nie, G. Stickney, and F.-G. Zeng, "Encoding frequency modulation to improve cochlear implant performance in noise," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 1, pp. 64–73, Jan. 2005.

[21] J. Laneau, J. Wouters, and M. Moonen, "Improved music perception with explicit pitch coding in cochlear implants," *Audiol. Neurotol.*, vol. 11, no. 1, pp. 38–52, 2006.

[22] M. Milczynski, J. Wouters, and A. van Wieringen, "Improved fundamental frequency coding in cochlear implant signal processing," *J. Acoust. Soc. Amer.*, vol. 125, no. 4, pp. 2260–2271, Apr. 2009.

[23] M. Milczynski, J. E. Chang, J. Wouters, and A. van Wieringen, "Perception of Mandarin Chinese with cochlear implants using enhanced temporal pitch cues," *Hearing Res.*, vol. 285, nos. 1–2, pp. 1–12, Mar. 2012.

[24] T. Francart, A. Osses, and J. Wouters, "Speech perception with F0mod, a cochlear implant pitch coding strategy," *Int. J. Audiol.*, vol. 54, no. 6, pp. 424–432, Jun. 2015.

[25] A. E. Vandali and R. J. M. van Hoesel, "Development of a temporal fundamental frequency coding strategy for cochlear implants," *J. Acoust. Soc. Amer.*, vol. 129, no. 6, pp. 4023–4036, Jun. 2011.

[26] A. E. Vandali, P. W. Dawson, and K. Arora, "Results using the OPAL strategy in Mandarin speaking cochlear implant recipients," *Int. J. Audiol.*, vol. 56, no. 2, pp. S74–S85, Oct. 2017.

[27] A. Vandali *et al.*, "Evaluation of the optimized pitch and language strategy in cochlear implant recipients," *Ear Hearing*, vol. 40, no. 3, pp. 555–567, 2019.

[28] L. Ping *et al.*, "Implementation and preliminary evaluation of 'C-tone': A novel algorithm to improve lexical tone recognition in Mandarin-speaking cochlear implant users," *Cochlear Implants Int.*, vol. 18, no. 5, pp. 240–249, Sep. 2017.

[29] J. N. Saba, H. Ali, and J. H. Hansen, "Formant priority channel selection for an 'n-of-m' sound processing strategy for cochlear implants," *J. Acoust. Soc. Amer.*, vol. 144, no. 6, pp. 3371–3380, 2018.

[30] X. Li, K. Nie, N. S. Imennov, J. T. Rubinstein, and L. E. Atlas, "Improved perception of music with a harmonic based algorithm for cochlear implants," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 21, no. 4, pp. 684–694, Jul. 2013.

[31] L. Geurts and J. Wouters, "Enhancing the speech envelope of continuous interleaved sampling processors for cochlear implants," *J. Acoust. Soc. Amer.*, vol. 105, no. 4, pp. 2476–2484, Apr. 1999.

[32] R. Koning and J. Wouters, "The potential of onset enhancement for increased speech intelligibility in auditory prostheses," *J. Acoust. Soc. Amer.*, vol. 132, no. 4, pp. 2569–2581, Oct. 2012.

[33] R. Koning and J. Wouters, "Speech onset enhancement improves intelligibility in adverse listening conditions for cochlear implant users," *Hearing Res.*, vol. 342, pp. 13–22, Dec. 2016.

[34] W. Nogueira, T. Rode, and A. Báchner, "Spectral contrast enhancement improves speech intelligibility in noise for cochlear implants," *J. Acoust. Soc. Amer.*, vol. 139, no. 2, pp. 728–739, Feb. 2016.

[35] W. K. Lai, N. Dillier, and M. Killian, "A neural excitability based coding strategy for cochlear implants," *J. Biomed. Sci. Eng.*, vol. 11, no. 7, pp. 159–181, 2018.

[36] S. Tabibi, A. Kegel, W. K. Lai, and N. Dillier, "A bio-inspired coding (BIC) strategy for cochlear implants," *Hearing Res.*, vol. 388, Mar. 2020, Art. no. 107885.

[37] E. A. Lopez-Poveda, A. Eustaquio-Martín, J. S. Stohl, R. D. Wolford, R. Schatzer, and B. S. Wilson, "A binaural cochlear implant sound coding strategy inspired by the contralateral medial olivocochlear reflex," *Ear Hearing*, vol. 37, no. 3, pp. e138–e148, 2016.

[38] E. A. Lopez-Poveda and A. Eustaquio-Martín, "Objective speech transmission improvements with a binaural cochlear implant sound-coding strategy inspired by the contralateral medial olivocochlear reflex," *J. Acoust. Soc. Amer.*, vol. 143, no. 4, pp. 2217–2231, Apr. 2018.
[39] E. A. Lopez-Poveda *et al.*, "Speech-in-Noise recognition with more realistic implementations of a binaural cochlear-implant sound coding strategy inspired by the medial olivocochlear reflex," *Ear Hearing*, vol. 41, no. 6, pp. 1492–1510, 2020.
[40] A. E. Vandali, C. Sucher, D. J. Tsang, C. M. McKay, J. W. D. Chew, and H. J. McDermott, "Pitch ranking ability of cochlear implant recipients: A comparison of sound-processing strategies," *J. Acoust. Soc. Amer.*, vol. 117, no. 5, pp. 3126–3138, May 2005.
[41] K. D. Morton, P. A. Torrione, C. S. Throckmorton, and L. M. Collins, "Mandarin Chinese tone identification in cochlear implants: Predictions from acoustic models," *Hearing Res.*, vol. 244, nos. 1–2, pp. 66–76, Oct. 2008.
[42] Q.-J. Fu, C.-J. Hsu, and M.-J. Horng, "Effects of speech processing strategy on Chinese tone recognition by nucleus-24 cochlear implant users," *Ear Hearing*, vol. 25, no. 5, pp. 501–508, Oct. 2004.
[43] X. Luo, Q.-J. Fu, C.-G. Wei, and K.-L. Cao, "Speech recognition and temporal amplitude modulation processing by Mandarin-speaking cochlear implant users," *Ear Hearing*, vol. 29, no. 6, p. 957, 2008.
[44] C.-F. Hwang, H.-C. Chen, C.-H. Yang, J.-P. Peng, and C.-H. Weng, "Comparison of Mandarin tone and speech perception between advanced combination encoder and continuous interleaved sampling speech-processing strategies in children," *Amer. J. Otolaryngol.*, vol. 33, no. 3, pp. 338–344, May 2012.
[45] X. Chen *et al.*, "Cochlear implants with fine structure processing improve speech and tone perception in mandarin-speaking adults," *Acta Oto-Laryngol.*, vol. 133, no. 7, pp. 733–738, Jul. 2013.
[46] B. Qi, Z. Liu, X. Gu, and B. Liu, "Speech recognition outcomes in mandarin-speaking cochlear implant users with fine structure processing," *Acta Oto-Laryngol.*, vol. 137, no. 3, pp. 286–292, Mar. 2017.
[47] D. Han *et al.*, "Lexical tone perception with HiResolution and HiResolution 120 sound-processing strategies in pediatric Mandarin-speaking cochlear implant users," *Ear Hearing*, vol. 30, no. 2, p. 169, 2009.
[48] R. Meddis, N. Clark, W. Lecluyse, and T. Jürgens, "Bioaid-ein biologisch inspiriertes Hörgerät," *Zeitschrift Audiologie/Audiol. Acoust.*, vol. 52, pp. 148–152, Dec. 2013.
[49] T. Järgens, N. R. Clark, W. Lecluyse, and R. Meddis, "Exploration of a physiologically-inspired hearing-aid algorithm using a computer model mimicking impaired hearing," *Int. J. Audiol.*, vol. 55, no. 6, pp. 346–357, Jun. 2016.
[50] F. Langner and T. Jürgens, "Forward-masked frequency selectivity improvements in simulated and actual cochlear implant users using a preprocessing algorithm," *Trends Hearing*, vol. 20, pp. 1–14, Jan. 2016.
[51] N. R. Clark, W. Lecluyse, and T. Járgens, "Analysis of compressive properties of the BioAid hearing aid algorithm," *Int. J. Audiol.*, vol. 57, no. 3, pp. S130–S138, May 2018.
[52] S. M. A. Ernst, S. Kortlang, G. Grimm, T. Bisitz, B. Kollmeier, and S. D. Ewert, "Binaural model-based dynamic-range compression," *Int. J. Audiol.*, vol. 57, no. 3, pp. S31–S42, May 2018.
[53] J. F. Patrick, P. A. Busby, and P. J. Gibson, "The development of the nucleus freedom cochlear implant system," *Trends Amplif.*, vol. 10, no. 4, pp. 175–200, 2006.
[54] P. P. Khing, B. A. Swanson, and E. Ambikairajah, "The effect of automatic gain control structure and release time on cochlear implant speech intelligibility," *PLoS ONE*, vol. 8, no. 11, Nov. 2013, Art. no. e82263.
[55] D. Wang and J. H. L. Hansen, "Speech enhancement for cochlear implant recipients," *J. Acoust. Soc. Amer.*, vol. 143, no. 4, pp. 2244–2254, Apr. 2018.
[56] R. V. Shannon, F.-G. Zeng, V. Kamath, J. Wygonski, and M. Ekelid, "Speech recognition with primarily temporal cues," *Science*, vol. 270, no. 5234, pp. 303–304, Oct. 1995.
[57] C.-M. Wu, K.-Y. H. Huang, and H.-C. Lin, "Effects of channel number, stimulation rate, and electroacoustic stimulation of cochlear implant simulation on Chinese speech recognition in noise," in *Proc. 7th Asia Pacific Symp. Cochlear Implants Rel. Sci.*, Singapore, 2009, p. RS2B-7.
[58] E. H.-H. Huang, C.-M. Wu, and H.-C. Lin, "Simulation of three auditory physiology based CI sound coding strategies with Mandarin speech," in *Proc. 12nd Asia Pacific Symp. Cochlear Implants Rel. Sci. (APSCI)*, Tokyo, Japan, 2019, p. D2-2.
[59] H.-M. Yang and J.-L. Wu, "Mandarin lexical neighborhood test (M-LNT) for pre-school children: Development of test and its validation," *J. Taiwan Otolaryngol. Head Neck Surg.*, vol. 40, no. 1, pp. 1–12, 2005.
[60] S. L. Nissen, R. W. Harris, and A. Dukes, "Word recognition materials for native speakers of Taiwan Mandarin," *Amer. J. Audiol.*, vol. 17, no. 1, pp. 68–79, Jun. 2008.
[61] S. L. Nissen, R. W. Harris, and K. B. Slade, "Development of speech reception threshold materials for speakers of Taiwan Mandarin," *Int. J. Audiol.*, vol. 46, no. 8, pp. 449–458, Aug. 2007.
[62] L. L. Wong, S. D. Soli, S. Liu, N. Han, and M.-W. Huang, "Development of the Mandarin hearing in noise test (MHINT)," *Ear Hearing*, vol. 28, no. 2, p. 70S–74S, 2007.
[63] E. A. Lopez-Poveda and R. Meddis, "A human nonlinear cochlear filterbank," *J. Acoust. Soc. Amer.*, vol. 110, no. 6, pp. 3107–3118, 2001.
[64] B. C. J. Moore, "Coding of sounds in the auditory system and its relevance to signal processing and coding in cochlear implants," *Otol. Neurotol.*, vol. 24, no. 2, pp. 243–254, Mar. 2003.
[65] N. Clark. (2012). *The Biologically Inspired Hearing Aid*. Accessed: Jul. 20, 2020. [Online]. Available: http://bioaid.org.U.K./
[66] L. Geurts and J. Wouters, "Coding of the fundamental frequency in continuous interleaved sampling processors for cochlear implants," *J. Acoust. Soc. Amer.*, vol. 109, no. 2, pp. 713–726, Feb. 2001.
[67] B. Swanson and H. Mauch, *Nucleus MATLAB Toolbox 4.20 Softw. User Manual*. Lane Cove, NSW, Australia: Cochlear, 2006.
[68] J. F. Santos, S. Cosentino, O. Hazrati, P. C. Loizou, and T. H. Falk, "Objective speech intelligibility measurement for cochlear implant users in complex listening environments," *Speech Commun.*, vol. 55, nos. 7–8, pp. 815–824, Sep. 2013.
[69] T. H. Falk *et al.*, "Objective quality and intelligibility prediction for users of assistive listening devices: Advantages and limitations of existing tools," *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 114–124, Mar. 2015.
[70] G. D. Watkins, B. A. Swanson, and G. J. Suaning, "An evaluation of output signal to noise ratio as a predictor of cochlear implant speech intelligibility," *Ear Hearing*, vol. 39, no. 5, pp. 958–968, 2018.
[71] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "An algorithm for intelligibility prediction of time-frequency weighted noisy speech," *IEEE Trans. Audio. Speech. Lang. Process.*, vol. 19, no. 7, pp. 2125–2136, Dec. 2011.
[72] H. Hu, M. E. Lutman, S. D. Ewert, G. Li, and S. Bleeck, "Sparse nonnegative matrix factorization strategy for cochlear implants," *Trends Hearing*, vol. 19, pp. 1–16, 2015.
[73] Y. H. Lai, F. Chen, S.-S. Wang, X. Lu, Y. Tsao, and C.-H. Lee, "A deep denoising autoencoder approach to improving the intelligibility of vocoded speech in cochlear implant simulation," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 7, pp. 1568–1578, Dec. 2017.
[74] F. Langner, A. Büchner, and W. Nogueira, "Evaluation of an adaptive dynamic compensation system in cochlear implant listeners," *Trends Hearing*, vol. 24, pp. 1–13, 2020.
[75] G. L. Mour ao, M. H. Costa, and S. Paul, "Speech intelligibility for cochlear implant users with the MMSE noise-reduction time-frequency mask," *Biomed. Signal Process. Control*, vol. 60, Dec. 2020, Art. no. 101982.
[76] R.-Y. Tseng, T.-W. Wang, S.-W. Fu, S.-W. Fu, C.-Y. Lee, and Y. Tsao, "A study of joint effect on denoising techniques and visual cues to improve speech intelligibility in cochlear implant simulation," *IEEE Trans. Cognit. Develop. Syst.*, early access, Aug. 17, 2020, doi: 10.1109/TCDS.2020.3017042.
[77] N. Y.-H. Wang *et al.*, "Improving the intelligibility of speech for simulated electric and acoustic stimulation using fully convolutional neural networks," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 184–195, 2021.
[78] P. C. Loizou and O. Poroy, "Minimum spectral contrast needed for vowel identification by normal hearing and cochlear implant listeners," *J. Acoust. Soc. Amer.*, vol. 110, no. 3, pp. 1619–1627, Sep. 2001.
[79] B. C. J. Moore, "Speech processing for the hearing-impaired: Successes, failures, and implications for speech mechanisms," *Speech Commun.*, vol. 41, no. 1, pp. 81–91, Aug. 2003.
[80] T. Green, A. Faulkner, and S. Rosen, "Spectral and temporal cues to pitch in noise-excited vocoder simulations of continuous-interleaved-sampling cochlear implants," *J. Acoust. Soc. Amer.*, vol. 112, no. 5, pp. 2155–2164, Nov. 2002.
[81] J. Laneau, M. Moonen, and J. Wouters, "Factors affecting the use of noise-band vocoders as acoustic models for pitch perception in cochlear implants," *J. Acoust. Soc. Amer.*, vol. 119, no. 1, pp. 491–506, Jan. 2006.