Pitch Perception With the Temporal Limits Encoder for Cochlear Implants

Huali Zhou, Alan Kan, *Member, IEEE*, Guangzheng Yu, Zhenyu Guo,
Nengheng Zheng, *Member, IEEE*, and Qinglin Meng

***Abstract*— The temporal-limits-encoder (TLE) strategy has been proposed to enhance the representation of temporal fine structure (TFS) in cochlear implants (CIs), which is vital for many aspects of sound perception but is typically discarded by most modern CI strategies. TLE works by computing an envelope modulator that is within the temporal pitch limits of CI electric hearing. This paper examines the TFS information encoded by TLE and evaluates the salience and usefulness of this information in CI users. Two experiments were conducted to compare pitch perception performance of TLE versus the widely-used Advanced Combinational Encoder (ACE) strategy. Experiment 1 investigated whether TLE processing improved pitch discrimination compared to ACE. Experiment 2 parametrically examined the effect of changing the lower frequency limit of the TLE modulator on pitch ranking. In both experiments, F0 difference limens were measured with synthetic harmonic complex tones using an adaptive procedure. Signal analysis of the outputs of TLE and ACE strategies showed that TLE introduces important temporal pitch cues that are not available with ACE. Results showed an improvement in pitch discrimination with TLE when the acoustic input had a lower F0 frequency. No significant effect of lower frequency limit was observed for pitch ranking, though a lower limit did tend to provide better outcomes. These results suggest that the envelope modulation introduced by TLE can improve pitch perception for CI listeners.**

***Index Terms*— Cochlear implant, pitch perception, temporal fine structure.**

I. INTRODUCTION

COCHLEAR implants (CIs) have been relatively successful in enabling most of their users to achieve good performance in speech perception in quiet [1]. However, CI users are still experiencing poor pitch perception, and thus are still struggling in other listening tasks such as speech-in-noise perception [2], music perception [3], speech intonation perception [4], and lexical tone perception in tonal languages [5]. In current CIs, input sounds are divided into sub-band channels. Each sub-band signal can be considered as a slowly varying envelope superimposed on a more fast oscillating carrier, i.e., the temporal fine structure (TFS) [6]. CIs encode the temporal envelope information contained in each channel as amplitude modulation changes on a train of electrical pulses.

Poor sensitivity to temporal modulation changes in electric hearing imposes constraints on effective delivery of TFS. Most CI users can only discriminate temporal changes in the range of approximately 50 to 300 Hz (though the upper limit may be higher in some CI users [7], [8], [9]) at individual electrode channels, which is often referred to as the temporal limit of pitch perception in CI users. More importantly, in most clinical CI strategies, only temporal envelopes from each channel is preserved, whereas the TFS is discarded. The lack of TFS may partly account for the aforementioned difficulties [10], [11].

TFS is an important acoustic cue for pitch perception [12], which plays a critical role in tasks that CI users find difficult. Given the importance of pitch, considerable efforts have been directed towards developing pitch enhancement algorithms for CI users, and some of them are now available in commercial CI devices. One approach has been designed to enhance periodicity by increasing modulation depth according to the instantaneous F0 in the channel envelopes. Examples are the F0mod strategy [13], [14], [15], [16], and the Optimized Pitch and Language (OPAL) strategy [17], [18], [19], [20]. Some other studies have recently explored new ways such as inserting pulses with short inter-pulse intervals (the SIPI strategy [21]). Another approach tried to enhance TFS by timing the electrical pulses to features of the acoustic signal. For example, the peak-derived timing (PDT) strategy [22] places a single pulse at the peak of the signal envelope, while the fine-structure processing (FSP) strategy [23], [24], [25], [26]

Manuscript received 26 April 2022; revised 26 July 2022 and 24 August 2022; accepted 28 August 2022. Date of publication 31 August 2022; date of current version 8 September 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 11704129 and Grant 61771320, in part by the Guangdong Basic and Applied Basic Research Foundation Grant under Grant 2020A1515010386 and Grant 2022A1515011361, and in part by the Science and Technology Program of Guangzhou under Grant 202102020944. *(Corresponding authors: Qinglin Meng; Nengheng Zheng.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethical Review Board of Shenzhen University.

Huali Zhou is with the Guangdong Key Laboratory of Intelligent Information Processing, College of Electronics and Information Engineering, Shenzhen University, Shenzhen, Guangdong 518060, China, and also with the Acoustics Laboratory, School of Physics and Optoelectronics, South China University of Technology, Guangzhou, Guangdong 510641, China.

Alan Kan is with the School of Engineering, Macquarie University, Sydney, NSW 2109, Australia.

Guangzheng Yu and Qinglin Meng are with the Acoustics Laboratory, School of Physics and Optoelectronics, South China University of Technology, Guangzhou, Guangdong 510641, China (e-mail: mengqinglin@scut.edu.cn).

Zhenyu Guo is with the School of Architecture, South China University of Technology, Guangzhou, Guangdong 510641, China.

Nengheng Zheng is with the Guangdong Key Laboratory of Intelligent Information Processing, College of Electronics and Information Engineering, Shenzhen University, Shenzhen, Guangdong 518060, China (e-mail: nhzheng@szu.edu.cn).

Digital Object Identifier 10.1109/TNSRE.2022.3203079

stimulates at every positive zero-crossing. The FSP strategy is commercially available in the devices of MED-EL (Innsbruck, Austria). However, mixed results have been reported with these strategies [27]. Though some of these strategies were reported to have substantial improvement, pitch perception of CI users is still unsatisfactory compared to that of normal-hearing listeners [28]. Hence, alternative methods for improving pitch perception for CI users are still worth investigating.

The Temporal Limits Encoder (TLE) strategy [29], [30] was designed to enhance the TFS representation in CIs. In contrast to prior attempts, the TLE strategy computes an envelope modulator by downshifting the mid-frequency channel information to a low-frequency range between the lower and upper temporal pitch limits. In this way, the rapidly-varying acoustic TFS, which is typically out of the perceptual limits of electric temporal pitch perception, is converted to a slowly-varying version that is within the perceptual pitch range of CI users. The TLE strategy has one free parameter, f_{lim}, which sets the lower temporal pitch limit. The TLE strategy has previously been tested in some pitch-related listening tests in normal-hearing (NH) listeners with offline vocoder simulations. The advantage of the TLE strategy over standard envelope-based strategies was observed in tasks including pure tone discrimination [29], binaural intelligibility level difference [31], Mandarin speech-in-noise reception, and Mandarin tone recognition [30]. However, the benefit of the TLE strategy for providing pitch perception benefits for CI users remains unclear because electric hearing is vastly different to acoustic hearing.

This paper presents a signal analysis of the potential pitch cues that are encoded by the TLE strategy and conducts listening experiments with CI users to evaluate the salience of these cues. For the listening experiments, a real-time version of the TLE strategy was developed and described in Section II. The real-time implementation used an overlapping frame-based approach with a Fast Fourier Transform (FFT)-based filterbank. This differed from the offline version of TLE used in the prior NH studies, which employed a filterbank of Butterworth filters. This change was necessary to allow a comparison of pitch perception performance between the TLE strategy and the Advanced Combinational Encoder (ACE) strategy [32]. ACE is a commonly used strategy for CI users with Cochlear Limited implants and it uses an FFT filterbank. Two listening tests were conducted with CI users to test the real-time version of the TLE strategy. Experiment 1 measured pitch discrimination thresholds for sounds encoded by TLE and ACE. In this experiment, f_{lim} was fixed at 50 Hz. Experiment 2 investigated the effect of the parameter f_{lim} on a pitch ranking (which sound is higher/lower in pitch) task. Compared to pitch discrimination, the pitch ranking task provides better insight into the usability of pitch information for real-world listening such as melody and tonal language perception.

II. THE TLE ALGORITHM

A. FFT-Based Real-Time Implementation

This section focuses on the FFT-based real-time implementation of the TLE strategy. This real-time implementation allows a direct comparison of performance in real CI users between both strategies.

Figure 1(a) shows the signal processing chain for both ACE and TLE. For each frame of sound, an FFT is applied to transform the acoustic input into the frequency domain. The FFT bins are then grouped into a number of frequency channels (typically equals to the number of electrodes). For ACE, the corresponding frequency bins in one channel are weighted and summed to provide a single magnitude value that is then used to amplitude modulate an electrical pulse (Fig. 1(b)). For TLE, a “*modulator*” is computed using a frequency down-shifting process (Fig. 1(c)) to amplitude modulate the electrical pulses [30], [31], [33].

The frequency down-shifting process in TLE is implemented for real-time testing as follows. First, for each channel, the FFT bins in each frame are transformed into the time domain using an Inverse Discrete Fourier Transform (IDFT). The IDFT of an N-point FFT $X_k (0 \le k \le N-1)$ is defined as

$$x_n = \frac{1}{N}\sum_{k=0}^{N-1} X_k e^{j\frac{2\pi}{N}kn}, 0 \le n \le N-1 \tag{1}$$

Here x is an N-point sequence in the time domain. When this is done frame by frame, we have an N-point IDFT sequence in each frame. In each frame at each CI channel, only one value is needed for each electrical pulse to control its pulse amplitude. In the case of ACE, the value is the magnitude of the weighted complex DFT outputs within corresponding channel. In TLE, we proposed to make use of the middle point selected from the N-point IDFT sequence, i.e., the point with $n = \frac{N}{2}$. For a channel that contains FFT bins X_p to X_q in a frame, the selected middle point is computed using

$$s[g] = \sum_{k=p}^{q} w_k X_k[g] e^{j\pi k} = \sum_{k=p}^{q} (-1)^k w_k X_k[g] \tag{2}$$

where w_k is the weight for the k^{th} FFT bin, and g is the frame number. With consecutive frames, s is a signal in the time domain for the channel.

Then, the frequency downshifting process is applied in the time domain. For a continuous-time band-limited channel signal, the frequency can be downshifted by multiplying with $e^{-j2\pi f_m t}$. This multiplication does not change the Hilbert envelope and spectral structure of the original signal. Only the spectral centroid is changed after the frequency-downshifting process. In frame-based real-time processing using an FFT filterbank, the frequency downshifting processing can be applied to the output of Eq. 2 using

$$\begin{aligned} v[g] &= s[g] e^{-j2\pi f_m (g-1) T_{shift}} \\ &= \sum_{k=p}^{q} (-1)^k w_k X_k[g] e^{-j2\pi f_m (g-1) T_{shift}} \end{aligned} \tag{3}$$

where g is the frame number and T_{shift} is the time shift between adjacent frames. After frequency downshifting, the real part of $v[g]$ is half-wave rectified to get the TLE modulator in the g_{th} frame.

The frequency of f_m in Eq. 3 should be set carefully because it determines the frequency of the downshifted signal. Recall the temporal pitch limits in electric hearing that most CI

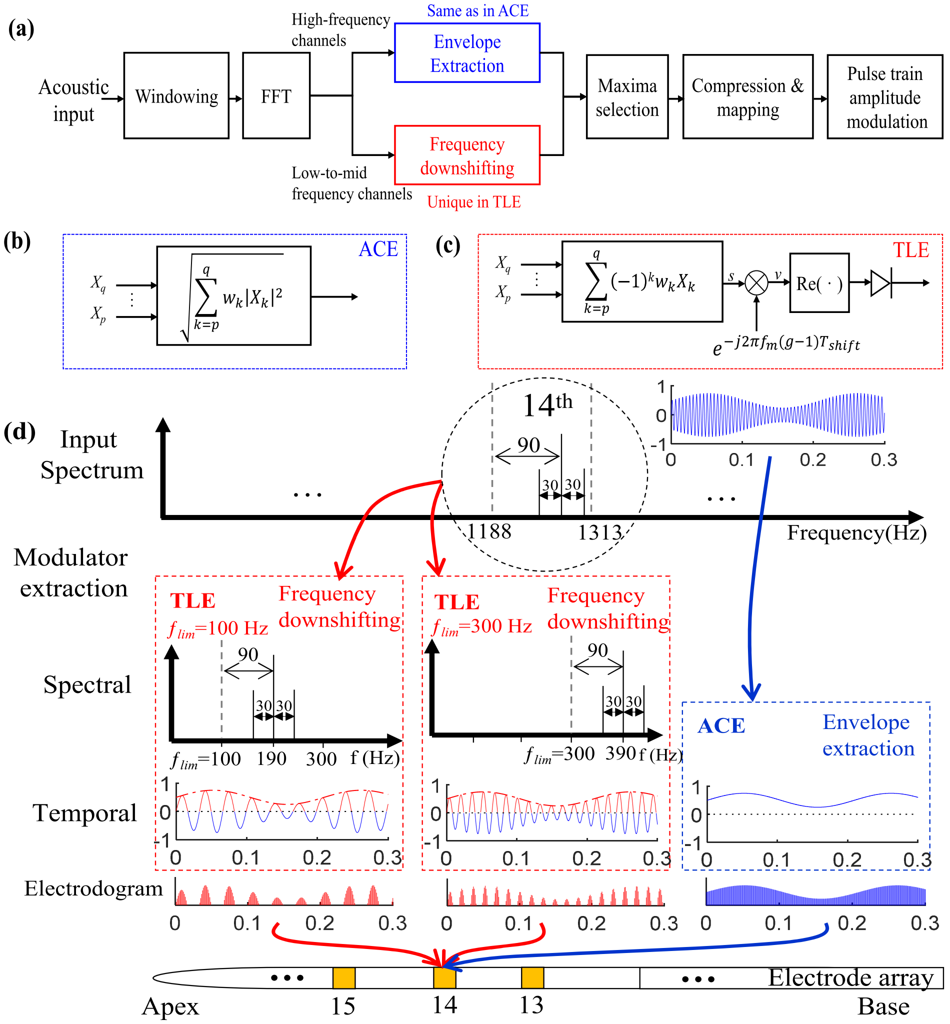

Fig. 1. Schematic diagrams of TLE and ACE. (a) Block diagram of the TLE strategy. The red rectangle illustrates the processing that is unique in TLE, and the rest processing is the same as in ACE. (b) Envelope extraction process based on FFT in ACE. (c) Frequency downshifting process based on FFT in TLE. (d) Schematic diagram of the difference in modulator extraction between TLE and ACE. For demonstration purpose, two different versions of TLE was included with f_{lim} set at different values, and the single-channel electrodograms are presented at a constant pulse rate of 800 pulses per second without compression, maxima selection and quantization into electric current levels.

users can only perceive temporal information from 50 Hz to 300 Hz [7], [8], [9]. These limits require TLE to have an intermediate frequency for f_m, so that the downshifted TFS could be within the limits. In TLE, f_m is set channel-specifically at

$$f_m = f_{low} - f_{lim}, \tag{4}$$

where f_{low} is the lower frequency bound of the channel and f_{lim} is a user-defined parameter representing the lower limit of temporal pitch perception on single electrode (typically higher than 50 Hz). After the frequency-downshifting process, f_{lim} is also the lower frequency bound of the downshifted signal.

Figure. 1(d) illustrates the difference in modulators between the two strategies for an artificial stimulus. In TLE, each frequency component is downshifted to a lower frequency while keeping the spectral structure. The only change is the spectral centroid, and the new centroid after frequency downshifting is determined by the user-defined parameter f_{lim}. For example, for the frequency component 90 Hz higher than the channel lower frequency bound in the 14^{th} channel (the dashed circle), after donwshifting its frequency is 90 Hz higher than f_{lim}. In the figure, for demonstration purpose, two versions of TLE were included, with f_{lim} = 100 and 300, respectively. Moreover, the spectral structure (two components 30 Hz lower and higher than the center component, respectively) is maintained after frequency downshifting.

The effect of the modulator lower limit f_{lim} can be seen from the examples in Fig.1(d). As mentioned before, it determines the lower frequency bound of the downshifted signal. The given input in this figure has different carrier frequencies after frequency downshifting of two different versions of TLE. The temporal waveform of the modulator of TLE in the right dashed rectangle is changing more fast than that in the left one. This is because f_{lim} was set at a higher value of 300 Hz in the right dashed rectangle, but at a lower value of 100 Hz in the left one. For a given original frequency, a higher f_{lim} results in a higher downshifted frequency.

Then, the downshifted subband signals are half-wave rectified and go through the same maxima selection and compression process as in the ACE strategy as the amplitude

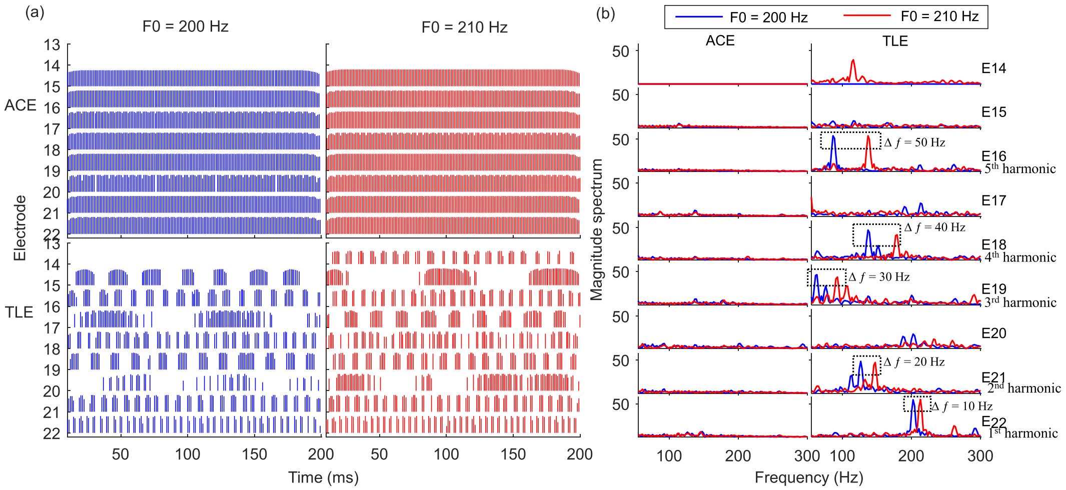

Fig. 2. Example electrodogram (a) and modulator spectrum at each electrode (b) of two 5-harmonic complex tones in ACE and TLE. For easier comparisons, modulator spectrum of the two consecutive complex tones were put together in the same figure in (b).

modulator of a pulse train with a constant pulse rate. Note that this within-channel frequency downshifting is done after bandpass filtering in channels with relatively narrow bandwidth. Therefore, in reference to ACE, only the modulators at several given (low-to-middle frequency) electrodes are changed, while the modulators in other channels with higher frequencies and wider bandwidths are not affected in TLE. The overall spatial pattern (frequency-electrode allocation) is kept the same as in regular strategies.

B. Signal Analysis

Two examples of different stimulation patterns between ACE and TLE (with f_{lim} = 50Hz) are shown in the stimulation patterns (also called electrodograms) in Fig. 2(a). The acoustic input was two complex tones with F0s of 200 and 210 Hz and each included the first 5 harmonics (equal amplitude among harmonics). It can be seen that the output modulated pulses of TLE are substantially different from those of ACE.

Figure. 2(b) shows the magnitude spectrum of each channels' pulse modulator extracted from the electrodogram for ACE (the left column) and TLE (the right column). In ACE, F0 information was quite limited in all electrodes, with no substantial peaks related to F0 in the magnitude spectrum. This is because of the narrow bandwidth (−3 dB bandwidth of 180 Hz [18]) in ACE. With this narrow bandwidth, little amplitude modulation can be expected for F0s of 180 Hz or higher in low frequency channels.

The TLE modulator spectrum is distinctly different from that of ACE. There were substantial spectral peaks related to F0 at electrodes that contained harmonic components, and the frequencies of these peaks increased with the increasing of F0. Further, the frequency difference between the original TFSs was maintained in the downshifted TFSs. For example, at E21 which contains the second harmonic component, the acoustic input has a 20 Hz difference in this range. After TLE processing, this 20-Hz difference is observed in the envelope modulators. This frequency difference is also preserved for the 3rd, 4th and 5th harmonics in E19, E18, and E16, respectively. Hence, it is reasonable to hypothesize that these differences in the envelope modulator created by TLE processing might help CI listeners to perceive a change in F0, when compared to ACE processing.

III. EXPERIMENT 1: PITCH DISCRIMINATION

A. Stimuli

Synthetic complex tones, comprised of the first five sinusoidal harmonic components of a fundamental frequency (F0) with a spectral roll-off of 10 dB per octave, were used as the acoustic stimuli. The starting phase of each harmonic component was randomized. Each complex tone was 400 ms in duration, gated on and off with a 20-ms sinusoidal ramp, and was successively presented with a 300-ms inter-stimulus interval. A three-interval, three-alternative forced-choice (3I-3AFC) procedure was used in this experiment. In each trial, two intervals were randomly assigned to have the same F0 value and the remaining interval had a different F0 value. Specifically, the F0s of complex tone pair in each trial were always centered around a center F0 value (denoted by F0c) with a difference denoted by ΔF0, namely, F0c + 0.5ΔF0 and F0c − 0.5ΔF0.

Four F0c values were tested in this experiment: 250, 313, 1000, and 1063 Hz, which are the center and upper cut-off frequency of electrodes 22 and 16 in the default 22-channel frequency allocation table of the Cochlear devices, respectively. 250 and 313 Hz are within the typical voice pitch range, while 1000 and 1063 Hz are in the upper musical pitch range.

B. Participants

Seven adult CI users (listed in Table I) participated in this study. All CI listeners used the clinical default frequency allocation tables (FAT) for 22 active electrodes, except for two listeners who used the default FAT for 20 active electrodes (C1 and C9, two deactivated electrodes at the apical end).

The real-time TLE and ACE strategies were implemented in MATLAB and presented to CI participants using a CCi-MOBILE research processor developed by University of

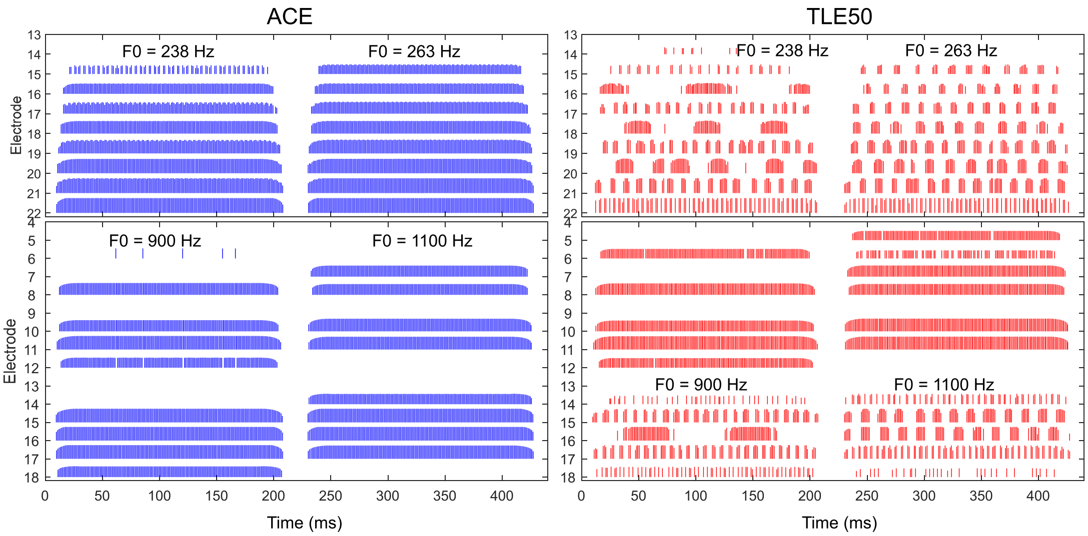

Fig. 3. Example stimulation patterns (electrodograms) for two stimulus pairs (one for each row, with F0c of 250 and 1000 Hz, respectively, and ΔF0 = F0c × 10%) processed by the two strategies. In each panel, the stimuli were a pair of complex tones with F0 of F0c−ΔF0/2 followed by F0c+ΔF0/2. Information regarding F0c and strategy is given in each electrodogram. The stimulus duration was reduced to 200 ms and the inter-stimulus interval was also reduced for illustration purposes.

Texas at Dallas [34], [35]. The output volume and microphone sensitivities were adjusted to compensate for potential loudness differences induced by modulator differences. This was done separately for the two strategies so that a speech-shaped noise presented at 65 dB(A) sounded equally loud with the two strategies. For this experiment, f_{lim} (the lower limit of the modulator) was set to 50 Hz. From here onwards, we refer to this condition as TLE50. In this way, electrodes 21 to 13 (21 to 15 for participants with 20 active electrodes) were downshifted to a lower frequency range above 50 Hz. The lowest frequency channel (electrode 22) was not downshifted but the original signal was used directly as the modulator without envelope extraction. Example electrodograms of TLE50 and ACE are shown in in Fig. 3. These electrodograms were generated using two stimulus pairs used in this experiment (described in Section III-A).

Nine normal-hearing (NH) listeners (three females and six males, aged 19 to 35 years, mean = 23.3 years) were recruited to provide baseline performance metrics for the task and experiment. NH listeners were students from South China University of Technology. They all reported normal hearing and no history of otologic pathology or neurological disorder.

All participants in both Experiment 1 and 2 received financial compensation for their participation. Written informed consent was obtained from all participants before the experiment, and all procedures were approved by the ethical review board at Shenzhen University.

C. Procedure

The experiment was conducted in a soundproof room with a background noise level less than 30 dBA. Stimuli were presented through an audio interface (Focusrite Scarlett 2i4) and a loudspeaker (Yamaha HS5) located 1 m in front of the listener at a level of approximately 65 dBA measured at the location of the center of the head when the listener was absent.

TABLE I
PARTICIPANT DEMOGRAPHIC AND DEVICE INFORMATION

ID	Gender	Age at testing	Etiology	Processor	CI experience (yr)	Active electrodes
C1	F	23	Congenital	CP810	19	20
C2	M	26	Drug induced	CP950	17	22
C9	F	36	Tympanitis	ESPrit 3G	9	20
C14^a	F	42	Drug induced	CP810	13	22
C16	F	25	Unknown	Freedom	3	22
C31	F	23	Sudden deafness	CP910	7	22
C27	M	38	Unknown	CP802	1	22
Mean	-	30.4	-	-	9.9	-

a: the only bilateral user, tested unilaterally with her preferred ear.

Participants were instructed to select the interval that sounded different in pitch from the other two on a computer screen. If unsure, participants were asked to make their best guess. A trial was scored as correct when the listener correctly identified the target interval containing the tone with F0 different from the other two. After response, visual feedback indicating whether the response was correct was provided trial by trial.

CI listeners were tested unilaterally. One CI listener (C27) had an acoustic hearing threshold of 60 dB HL at 125 Hz and 75 dB HL at 250 Hz in the contra-lateral ear. This listener was tested with the contralateral ear plugged and muffed. None of the other CI listeners had residual acoustic hearing in the implanted ear or the ear contralateral to the implant. In contrast, NH listeners performed the task while listening with both ears. While the testing conditions are not the same, this testing reflects the real listening performance for both groups. In most cases, people with normal hearing listen with two ears, whereas most CI users are implanted unilaterally in China [36].

Fundamental frequency difference limens (F0DLs) were measured using an adaptive two-down one-up (2D1U) procedure that tracks the 70.7% correct point on the psychometric function [37]. The procedure adapts ΔF0 to determine the minimum ΔF0 that could be discriminated. The initial ΔF0 was set at a relatively large value of 50% of F0c to ensure that the F0 difference could be discriminated easily. The ΔF0 was increased by a factor after each incorrect response, and decreased by a factor after two consecutive correct responses. This factor was initially 2.0. It was reduced to 1.6 after the second reversal, 1.3 after the fourth reversal, 1.2 after the sixth reversal, and then 1.1 after the eighth reversal. The adaptive procedure terminated after 13 reversals with F0DL based on the geometric mean of the last eight reversals.

For each CI listener, there were four F0c conditions (250, 313, 1000, 1063 Hz) and two strategies (ACE and TLE50). Each F0c and strategy combination was tested twice in a random order. Thus, each CI listener completed 16 adaptive tracks in total. For NH listeners, the strategy condition was not applicable. Hence, each NH listener completed two runs for each of the four F0c, i.e, 8 adaptive tracks in total. The geometric mean result of the two runs was used as the final F0DL for each F0c tested. Prior to the formal tests, a training session (∼30 minutes) was conducted to familiarize listeners with the research interface and the test procedure.

D. Results

Group results from CI participants are shown in Fig. 4(a) along with the group results from NH listeners. Compared with the NH group, much higher F0DLs (poorer pitch discrimination ability) were observed for the CI group. Specifically, the mean F0DLs of the NH group at the four F0c were mostly between 1% and 2%, whereas those of the CI group fell between 10% and 30%, which is higher than those of the NH group by an order of magnitude. The mean F0DL of the CI group at F0c from 250 to 1063 Hz was 27.7%, 27.4%, 10.9%, and 12.2% for ACE, and 18.2%, 13.8%, 12.7%, and 10.7% for TLE50.

To examine the effects of strategy and F0c on F0DL, a repeated-measures two-way analysis of variance (ANOVA) was conducted. Measured F0DL data were analyzed in logarithmic space (see [38] for a rationale for applying a logarithmic transformation on the thresholds prior to statistical analyses). Data normality was confirmed by the Shapiro-Wilk test. Results of the Mauchly's Test of Sphericity showed that sphericity was not violated. Significant effects of strategy ($F_{1,6} = 7.924$, $p = 0.031$) and F0c ($F_{3,18} = 4.935$, $p = 0.011$) were observed, and also a marginal effect of interaction between strategy and F0c ($F_{3,18} = 3.112$, $p = 0.052$). Pairwise comparisons with Bonferroni corrections revealed that TLE50 had significantly better performance compared to ACE for F0c of 313 Hz [mean difference (md) = 13.7%, $p = 0.013$], and not for F0c of 250 ($md = 9.5\%$, $p = 0.180$), 1000 ($md = -1.85\%$, $p = 0.473$), and 1063 Hz ($md = 1.5\%$, $p = 0.158$).

Figure 4(b) depicts individual differences in F0DLs when listening with TLE50 vs ACE. At the 250-Hz F0c, two of the participants performed worse with TLE (negative improvement), and the rest five had considerable improvement with TLE. At the 300-Hz F0c, all the participants performed better with TLE. At the 1000-Hz F0c, no participant had distinct improvement with TLE. At the 1063-Hz F0c, most participants had no distinct improvement with TLE either.

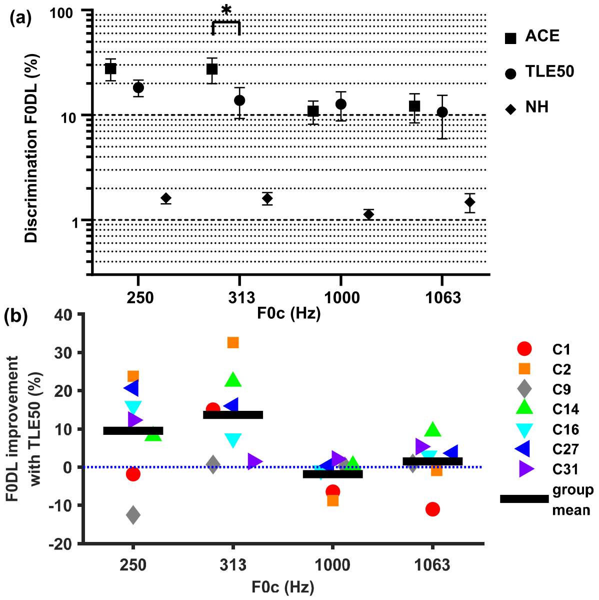

Fig. 4. Pitch discrimination results. (a):Mean discrimination F0DLs of the CI group using the two strategies (i.e., ACE and TLE50) and the NH group. Lower values indicate better performance. TLE50 refers to the TLE strategy with a modulator lower limit of 50 Hz. Error bars indicate the standard error of mean (error bars shorter than the size of symbols were not shown) and the asterisk denotes the statistical significance ($p < 0.05$). (b):Improvement in F0DL when using TLE50 compared with ACE (calculated by subtracting TLE50 F0DL from that of ACE). A positive value indicates that TLE had better performance than ACE. Each symbol denotes the difference in F0DL for an individual CI participant, and the thick black line denotes the group mean.

E. Discussion

The F0DLs measured in this study indicate that CI users' F0DLs were substantially poorer than those in NH listeners (mean across all conditions of 19.5%, 13.9% and 1.5% for ACE, TLE50, and NH, respectively). The variability in F0DLs was substantial across CI users, with F0DL ranges of 3.0% to 68.9% and 2.1% to 36.3% for ACE and TLE50, respectively. These results are consistent with previous reports with similar measure in the literature. Previous studies report an F0DL of 0.5% to 2% in NH listeners [39], [40], [41], and an F0DL which is an order of magnitude higher in CI users [42], [43], [44], [45], [46]. For instance, Goldsworthy [42] reported F0DLs at standard F0s of 110, 220 and 440 Hz using band-pass filtered synthetic harmonic complex tones in CI and NH listeners. In that study, 6 out of 9 CI users were listening with the ACE strategy, 2 out of 9 listened with the HiRes strategy and the remaining participant used the SPEAK. Mean F0DL of CI vs NH was 12.5% vs 1.4%, with an F0DL range of 2.6 to 28.5% for the CI group. Marx *et al.* [45] compared F0DLs of a CI group and an NH group at standard F0s of 110, 220, 400, 500, and 750 Hz. Results of each F0 were not reported in that study, but averaged across standard F0s. The F0DLs of the CI group in that study ranged between 7 to 37%, with a mean of 34%, compared to 2.2% in the NH group.

The group mean benefit for TLE50 was 9.5 and 13.7 percentage points for F0c of 250 and 313 Hz, respectively. The mean benefit of 9.5% (1.3 semitones) for F0c = 250 Hz is relatively large. However, it was not found to be statistically significant. All participants demonstrated a trend of higher mean scores from TLE50 except for C9 and C1 at 250 Hz. Note that these two participants are the ones who had two deactivated electrodes. Hence, their frequency allocations were different from the other participants who had all electrodes active. For these listeners, TLE coding was applied to less electrodes than other listeners (the electrodes that had frequency downshifting were electrodes 21 to 13 in the default map, but were 21 to 15 in map with 20 active electrodes). These differences, and any subsequent effect of the differences with maxima selection, might explain the lack of benefit for these listeners at 250 Hz.

No benefit was observed for TLE50 at high F0c (1000 and 1063 Hz). One possible explanation for the lack of benefit is that the frequency downshifting process in the TLE strategy was only applied in one electrode channel. For higher F0c (around 1000 Hz), only the lowest one component in the 5-harmonic stimulus falls within the 300 to 1500 Hz range in which the TLE strategy was applied (see Fig 3). Hence, the outputs of both TLE and ACE differed by only one harmonic component which may not be a salient enough cue to improve pitch discrimination. Taken together with the results of the two participants who had deactivated electrodes, the results suggest that it might be necessary to apply TLE processing to all electrode channels in order to maximize pitch discrimination benefits. The cues that led to the observed benefit of TLE are discussed in Section V-B.

IV. EXPERIMENT 2: PITCH RANKING

A. Rationale

Experiment 1 showed that the TLE strategy has the potential to improve pitch discrimination performance in CI users. However, real-world listening (such as melody perception) requires both the ability to detect a pitch change as well as the ability to tell the direction of the pitch change. In this experiment, a pitch ranking task was used in which the participant was required to pick the sound that was higher in pitch. Further, the effect of the modulator lower limit, f_{lim}, was examined.

B. Participants

Eleven adult CI users (listed in Table II) and ten NH participants (five females, aged 19 to 28 years, mean = 25 years) participated in this experiment.

C. Stimuli and Procedure

The pitch ranking experiment was conducted in the same soundproof room as Experiment 1. The same synthetic complex-tone pairs were used as the test stimuli. A 2I-2AFC task was used for this experiment. Participants were instructed to choose the interval with higher pitch, and a trial was scored as correct when the participant correctly identified the target interval containing the higher-frequency tone. Training was done before the formal tests following a similar procedure as in Experiment 1. No feedback was given during testing. Only tone pairs around F0c of 250 and 313 Hz were tested in this experiment, which are also in the range of typical voice F0s.

TABLE II
PARTICIPANT DEMOGRAPHICS IN EXPERIMENT 2

ID	Gender	Age at testing	Etiology	Clinical device	CI experience (yr)	Active electrodes
C9	F	36	Tympanitis	ESPrit 3G	9	20
C14[a]	F	42	Drug induced	CP810	13	22
C16	F	25	Unknown	Freedom	3	22
C27	M	38	Unknown	CP802	1	22
C31	F	23	Sudden deafness	CP910	7	22
C32	M	62	Sudden deafness	CP910	1	22
C33	M	31	Sudden deafness	Freedom	8	22
C34	M	46	Unknown	CP802	1	22
C26	M	22	Meningitis	Freedom	13	22
C36	M	26	Unknown	CP950	1	22
C37	M	44	Tympanitis	Freedom	8	22
Mean	-	35.9	-	-	5.9	-

[a]: the only bilateral user, tested unilaterally with her preferred ear.

To reduce testing time and alleviate participant fatigue, the 2D1U adaptive procedure was adapted in a way similar as that used in [47]. It terminates after 13 reversals or when the standard deviation (SD) of six consecutive reversals gets within twice of the SDs collected in Experiment 1 (i.e., SD = 9.77 Hz at F0c = 250 Hz, and SD = 12.65 Hz at F0c = 313 Hz). The buffer consecutive reversals to be used for SD calculation was cleared after six consecutive correct or wrong responses to make sure the procedure always terminates at a convergence level. The geometric mean of the last six reversals was calculated as the pitch-ranking F0DL.

The same strategy implementation and fitting as described in Section III-B was used in this experiment. For this experiment, f_{lim} was set to 100, 200 or 300 Hz (denoted by TLE100, TLE200 and TLE300, respectively). Together with the ACE strategy, there were four strategy conditions, and they were all tested with the CCi-MOBILE. The electrical stimulation patterns (electrodograms) of the four strategy conditions to the same input are illustrated in Fig. 5. The input was a consecutive pair of harmonic tones described in Section III-A.

Three test blocks, each including 8 test conditions (4 strategies × 2 reference F0s) in separate runs, were administered. The order of test conditions was randomized for each block and participant. For each test condition, the geometric mean of the three measured F0DLs were calculated as the final F0DL. Each NH participant completed three runs for each of the two reference F0s using the same stimuli and procedure as for the CI participants.

D. Results

Pitch ranking performance of both the NH and CI group using the four strategy conditions (ACE, TLE100, TLE200 and TLE300) at the two F0c (250 and 313 Hz) are shown in Fig. 6. Similar to Experiment 1, the CI group had much higher F0DLs than the NH group, by about an order of magnitude. Performance between the two F0cs were similar within the CI group. The mean F0DL of the CI group at the 250-Hz

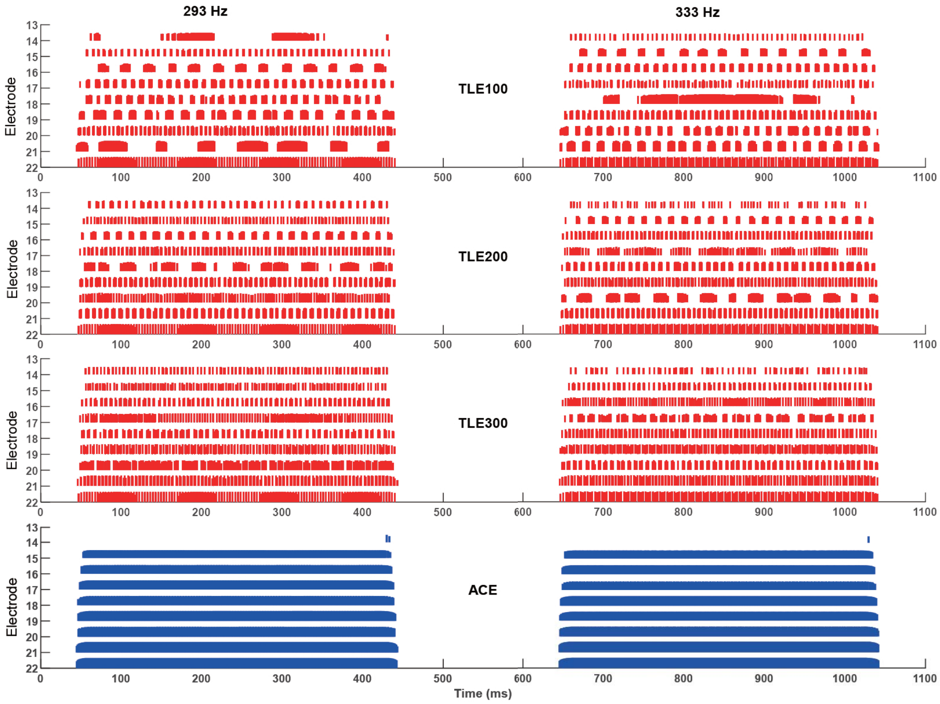

Fig. 5. Example electrodograms of TLE100, TLE200, TLE300, and ACE. The acoustic stimulus was the same as the top two panels in Fig. 3 (F0c = 250 Hz, ΔF0 = 25 Hz).

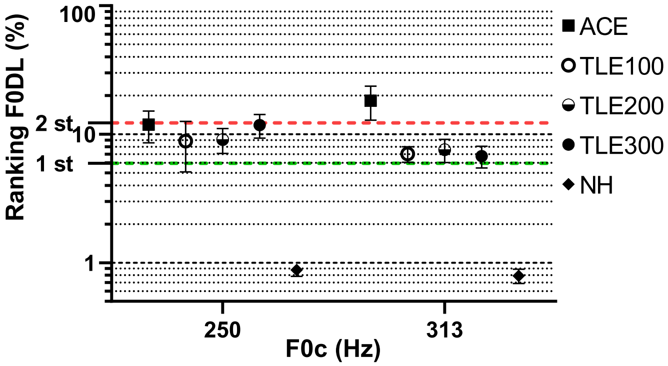

Fig. 6. Mean F0DLs of the CI group with four strategy conditions and the NH group. Lower values indicate better performance. Green and red dashed lines denote the levels of one semitone (1 st) and two semitones (2 st), respectively.

F0c was 11.9%, 8.8%, 9.1%, and 11.8% for ACE, TLE100, TLE200, and TLE300, respectively, and at the 313 Hz F0c it was 18.3%, 7.1%, 7.6%, and 6.8%, respectively.

A repeated-measures two-way ANOVA was conducted to understand the effects of strategy, F0c, and the interaction between strategy and F0c on the pitch-ranking F0DL. There was a significant main effect of strategy ($F_{3,30} = 4.586$, $p = 0.009$), but not of F0c ($F_{1,10} = 0.017$, $p = 0.900$).There was no significant interaction between the two factors ($F_{3,30} = 2.626$, $p = 0.069$).

As there was no significant difference between the two F0c and no significant interaction effect, F0DLs of the two F0c were pooled together to analyze the difference among strategies. F0DLs of F0c of 250 and 313 Hz were averaged for each strategy, and results are shown in Fig. 7. The group mean F0DL of ACE, TLE100, TLE200, and TLE300 was 15.1%, 7.9%, 8.3%, and 9.3% respectively. Multiple comparisons with Bonferroni corrections revealed that the mean F0DL of TLE100 was significantly lower (better) than that obtained with the ACE strategy ($p = 0.006$). The difference between TLE200 and ACE, and between TLE300 and ACE, were not significant ($p = 0.062$ and $p = 0.188$, respectively). The differences among TLE100, TLE200,and TLE300 were not significant either.

E. Discussion

The F0DLs measured in this experiment are consistent with previous pitch ranking studies using similar complex-tone stimuli in the literature. For NH listeners, Qin and Oxenham [40] reported pitch ranking F0DLs of ~0.8% measured at a nominal F0 of 220 Hz, while the NH F0DLs measured in this study are 0.9% and 0.8% at F0c of 250 and 313 Hz, respectively. Jiam *et al.* [48] reported pitch ranking F0DLs of NH listeners for a reference F0 roved between 100 and 150 Hz. NH listeners randomly assigned to two groups in that study showed F0DLs of ~0.7% (0.125 semitones) and ~2.9% (0.5 semitones) before the designed music training. For CI users, the group mean F0DL using the ACE strategy in this study was 15.1% at F0c of 250 and 313 Hz. Kang *et al.* [49] reported pitch ranking thresholds for a large group of CI users (42 CI users, most using ACE strategy). The mean threshold in that study was 18.2% (2.9 semitones) and 21.7% (3.4 semitones) for base frequency of 263 and 330 Hz, respectively. Kang *et al.* [48] also reported F0DLs of 21 CI listeners with various strategies including ACE randomly assigned to two groups. The baseline F0DL before the designed music training was ~9.1% (1.5 semitones) and ~18.9% (3 semitones) for the two CI groups in that study.

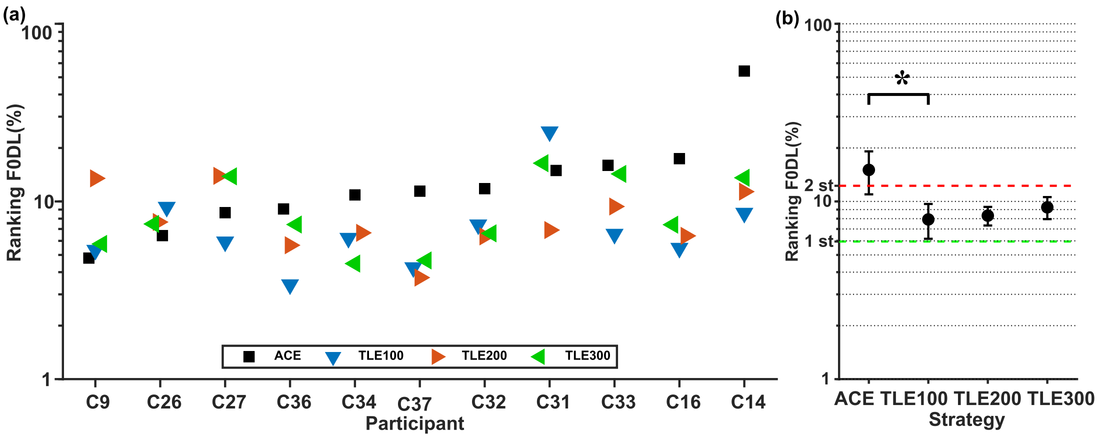

Fig. 7. Averaged F0DLs of the two F0c. (a) shows the individual results, and (b) shows the group means and standard error of mean. Lower green and upper red dashed lines denote the levels of one semitone (1 st) and two semitones (2 st), respectively. The asterisk denotes a significant difference ($p < 0.05$).

Our result are comparable with those in that study, especially the second group. Vandali *et al.* [50] investigated the effect of training on F0DLs in two groups (N = 5 for each group) of CI users using mostly the ACE strategy. The control group who did not take the training program had group mean F0DLs of 18.1%, 21.1%, and 16.2% across sessions 1 to 3 (with a time interval between sessions of at least eight weeks), respectively.

While Experiment 1 found a benefit in the TLE strategy in detecting a pitch change, results in Experiment 2 suggested that the TLE strategy can also improve CI users' perception of the direction of a pitch change. The different patterns among channels of the TLE strategy compared to those of ACE strategy (see Fig. 3) intuitively support the benefit in pitch discrimination. However, no such straightforward representation for the improvement in pitch ranking can be observed in the electrodograms in Fig. 5. For the higher F0, envelope fluctuation rates in all the electrodes are not consistently higher than those of the lower F0. In some electrodes, the envelope fluctuation rates can even be lower for a higher F0. This is because 1) the FFT-based bandpass filters had non-ideal attenuation outside the frequency ranges specified in the frequency allocation table, thus leading to frequency leakage from undesirable frequencies, and 2) when a harmonic component increases (or decreases) from one channel into a neighboring higher (or lower) channel, the amplitude modulation will become slower (or faster). Nonetheless, the experiment showed positive results, which suggests that the TLE strategy can introduce downshifted within-channel TFS cues in the form of variations in the envelope fluctuation which enhances a CI user's ability to determine the direction of pitch change.

Regarding the effect of the modulator lower limit, f_{lim}, results in this experiment show that TLE100 outperformed ACE in pitch ranking, whereas TLE200 and TLE300 did not. The best performance of TLE100 indicates that the downshifted within-channel TFS cues with 100-Hz f_{lim} are sufficiently salient to provide a substantial benefit over ACE. For TLE200 and TLE300, enhancement of downshifted within-channel TFS cues was not substantial enough to provide a significant benefit compared to ACE. In theory, a lower f_{lim} should produce a more salient temporal pitch than a higher f_{lim}. While the predicted trend can be observed (Fig. 7), there was no significant difference among the three TLE conditions. This may, perhaps, be due to the limited number of participants and their individual ability to make use of the downshifted within-channel TFS cues. The lack of significant difference among the three TLE conditions suggest that the choice of f_{lim} may not be critical for pitch ranking, although a lower f_{lim} could have a small advantage.

V. GENERAL DISCUSSION

This study investigated whether the FFT-based real-time TLE strategy can improve pitch perception of CI users with two experiments. Results showed that difference limens improved with TLE when a harmonic complex was completely encoded by electrode channels that used TLE processing. Further, the choice of f_{lim} does not seem to correlate with performance, though a lower f_{lim} could potentially provide a small advantage.

A. Improvement of CI Pitch Perception

Improving pitch perception of CI users has been a challenging task, for which some strategies have made some encouraging achievements. For example, compared to ACE, the OPAL strategy has been reported to have a 8.5% intonation-perception benefit [20], a 15% pitch-ranking benefit of sung-vowel stimuli [18], and a 6% lexical tone recognition benefit [19]. The F0mod strategy has been shown to have significant improvement compared to ACE in pitch ranking of synthetic harmonic complexes [14], [16], in F0 discrimination of musical tones [13], in melodic contour identification and familiar melody identification [14], and in Mandarin lexical tone recognition [15].

However, the temporal pitch limits in electric hearing impose great constraints on performance of the above mentioned strategies. The sensation of amplitude modulation in electric hearing is limited to the temporal pitch limit, namely, 300 Hz in most CI users. Therefore, it is difficult for these strategies to break through the temporal pitch limit for F0s close to 300 Hz. For example, Milczynski *et al.* [15] reported significant better Mandarin lexical tone recognition with the F0mod strategy than with ACE for the male voice (F0 ranged from 81 to 195 Hz), but not for the female voice (F0 ranged from 120 to 325 Hz). The authors also stated that tone recognition is likely to be particularly challenging for the female voice as the F0 contours for females approach the limits of effective temporal pitch perception in electric hearing. An F0-discrimination benefit of F0mod was observed at reference F0s of 130.8 and 185 Hz, but not at the reference F0 of 370 Hz [13]. In [14], a pitch-ranking benefit of F0mod was found at reference F0s of 131 and 165 Hz, but not at reference F0s of 208 and 262 Hz (comparable to the frequency range in our study). The scores of both F0mod and ACE were close to chance level at the reference F0 of 262 Hz even with an F0 difference of 4 semitones (262-330 Hz) in that study. The effect of TLE observed in this study brought new insight to the challenge of improving pitch perception for F0s close to 300 Hz. Due to the limited stimuli used in this study, future work evaluating pitch perception with TLE in a broader range of stimuli is warranted.

The observed benefit in this study suggests that the transposed TFS in TLE is accessible to CI users, and can be used for pitch perception. Although some participants reported that sounds are lower in pitch with TLE vs ACE (similar as related fine structure strategies, e.g., [51], [52]), prior experiments have shown that the TLE strategy has comparable speech recognition performance [33]. It is likely that given sufficient experience with TLE, participants may get used to sound of TLE and learn to utilize the temporal information provided by TLE. TLE was tested acutely (without prolonged exposure in this study) and was compared with ACE which all CI listeners have been using for years. It usually takes several weeks or months for CI users to adapt to a new sound coding strategy [53], [54]. Therefore, the obtained results can be considered promising and improved pitch perception performance is likely if listeners are provided with a more extensive adaptation period.

B. Model Analysis: Place Cues or Temporal Cues?

As both place (spectral) cues and temporal cues may be different between two complex tones with different F0s, it is important to determine the role of place and temporal cues to the observed outcomes, especially at the lower center F0s (i.e., 250 and 313 Hz) where effects of TLE were observed.

To quantify the place cues in electrodograms provided by both strategies, the mean place-centroid of stimulation (the gravity center of electric stimulation) was calculated following a similar method used in [55] and [56]. Specifically, it was calculated using Eq.(5).

$$C = \frac{\sum_{e=1}^{22} \sum_{i=1}^{N} e \times m(e,i)}{\sum_{e=1}^{22} \sum_{i=1}^{N} m(e,i)} \tag{5}$$

where e denotes the electrode number, N is the total frame number, and $m(e,i)$ denotes the current level of the electric pulse on electrode e in frame i. The distance between the centroids of the excitation patterns of two sounds determines the discriminability of the two sounds based on place cues.

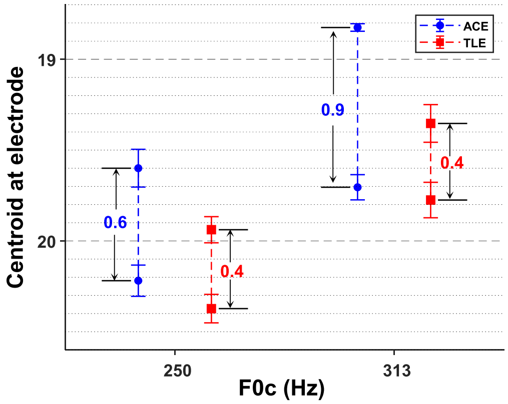

Fig. 8. Mean centroids with the ΔF0 set at the F0DL measured for both strategy. Error bars indicate the standard variations.

Figure 8 depicts the distributions of centroids of two stimuli with different F0s (lower points: F0 = F0c − F0DL/2; upper points: F0 = F0c + F0DL/2) at each F0c (left: 250 Hz; right: 313 Hz) for each strategy (red: ACE; blue: TLE50). Here, F0DL represents the mean difference limens accquired in Experiment 1. Note that the two stimuli F0DLs at each F0c were different for the two strategies (i.e., 27.2% (ACE) vs 18.2% (TLE), and 27.4% (ACE) vs 13.8% (TLE) for F0c of 250 and 313 Hz, respectively). Each distribution (represented by the mean and error bar) was obtained through 100 simulations. In each simulation, a complex tone with a roved overall intensity and randomized component initial phases as used in the experiments was generated. Then, the strategy processing was applied to get the electrodogram, and a spectral centroid was calculated using Eq.(5). The centroid distance between the lower and upper mean centroids was labeled in the number besides each line segment.

Model analysis results showed that participants were able to discriminate complex tones with a smaller centroid distance with TLE compared to ACE (shown in Fig. 8). This suggests that the observed benefit of TLE was not a result of place cues (i.e., larger centroid distance), and that participants were using some cue(s) other than the place cues to aid pitch perception with TLE.

Further, centroid distances were measured with a broader range of conditions than those tested in the experiments. In specific, centroid difference was also calculated as a function of F0c and ΔF0 using the same simulation method as used in Fig. 8. Results are shown in Fig. 9(a) and Fig. 9(b), respectively. Figure 9(a) presents ΔC for the two low F0c (i.e., 250 and 313 Hz) with ΔF0 from 1 to 5 semitones, and Fig. 9(b) presents ΔC for F0c from 100 to 313 Hz in a step of 50 Hz with ΔF0 of 3 and 4 semitones. In both panels, the curves of TLE were close to or slightly lower than those of ACE, suggesting that TLE provided equivalent or slightly less place information compared to ACE.

Based on the above analysis on the place cues, it is concluded that TLE provides equivalent or slightly less place information compared to ACE for the stimuli used in this

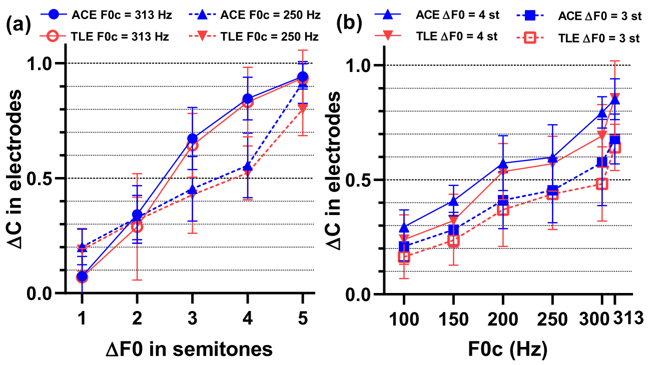

Fig. 9. Centroid distance as a function of ΔF0 (a), and as a function of F0c (b). st stands for semitones and error bars indicate the standard deviations.

study. The observed performance difference between the two strategies is therefore not likely to be affected by the place cues, but presumably to be affected by the temporal cue.

One potential limitation with the TLE strategy is that the harmonic relationship of components in the original acoustic stimulus may not be maintained after frequency transposition. This could be caused by the arbitrary frequency allocation and frequency downshifting. A frequency increasing or decreasing trend within a channel can be well-preserved after frequency downshifting. However, when the change in frequency crosses the upper or lower limits of one channel to another, the change is not preserved with TLE processing. However, in this case, the change in the stimulating electrode might be able to compensate via place pitch. For higher F0 (>∼180 Hz) perception, ACE could rely only on the place cues, while TLE could also use the place cues and at the same time introduces novel temporal cues (i.e., the slowly-varying within channel TFS).

VI. CONCLUSION

TLE and ACE strategies provided equivalent place cues but different temporal modulation cues for complex-tone stimuli. Listening tests showed a 9.5 percentage point improvement in pitch discrimination results and a 7.2 percentage point improvement for pitch ranking when listening with TLE compared to ACE. The choice of modulator lower limit f_{lim} does not seem to affect performance, though a lower f_{lim} seems to provide a small improvement in overall pitch ranking performance. These findings suggest that TLE may be a promising approach to improving pitch perception in CI users.

ACKNOWLEDGMENT

The authors would like to thank all the research volunteers that generously donated their time to participate in this study.

REFERENCES

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

[2] R. A. Chiea, M. H. Costa, and J. A. Cordioli, "An optimal envelope-based noise reduction method for cochlear implants: An upper bound performance investigation," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 29, pp. 1729–1739, 2021.

[3] W. Nogueira, A. Nagathil, and R. Martin, "Making music more accessible for cochlear implant listeners: Recent developments," *IEEE Signal Process. Mag.*, vol. 36, no. 1, pp. 115–127, Jan. 2019.

[4] M. Chatterjee *et al.*, "Acoustics of emotional prosody produced by prelingually deaf children with cochlear implants," *Frontiers Psychol.*, vol. 10, p. 2190, Sep. 2019.

[5] Q. Meng, N. Zheng, and X. Li, "Loudness contour can influence Mandarin tone recognition: Vocoder simulation and cochlear implants," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 6, pp. 641–649, Jun. 2017.

[6] S. Rosen, "Temporal information in speech: Acoustic, auditory and linguistic aspects," *Philos. Trans. Roy. Soc. London B, Biol. Sci.*, vol. 336, no. 1278, pp. 367–373, 1992.

[7] B. C. Moore and R. P. Carlyon, "Perception of pitch by people with cochlear hearing loss and by cochlear implant users," in *Pitch*. New York, NY, USA: Springer, 2005, pp. 234–277.

[8] Y.-Y. Kong and R. P. Carlyon, "Temporal pitch perception at high rates in cochlear implants," *J. Acoust. Soc. Amer.*, vol. 127, no. 5, pp. 3114–3123, May 2010.

[9] N. Zhou, J. Mathews, and L. Dong, "Pulse-rate discrimination deficit in cochlear implant users: Is the upper limit of pitch peripheral or central?" *Hearing Res.*, vol. 371, pp. 1–10, Jan. 2019.

[10] H. Jones, A. Kan, and R. Y. Litovsky, "Comparing sound localization deficits in bilateral cochlear-implant users and vocoder simulations with normal-hearing listeners," *Trends Hearing*, vol. 18, 2014, Art. no. 2331216514554574.

[11] A. Ihlefeld and R. Y. Litovsky, "Interaural level differences do not suffice for restoring spatial release from masking in simulated cochlear implant listening," *PLoS ONE*, vol. 7, no. 9, Sep. 2012, Art. no. e45296.

[12] Z. M. Smith, B. Delgutte, and A. J. Oxenham, "Chimaeric sounds reveal dichotomies in auditory perception," *Nature*, vol. 416, no. 6876, pp. 87–90, Mar. 2002.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] A. E. Vandali and R. J. M. van Hoesel, "Enhancement of temporal cues to pitch in cochlear implants: Effects on pitch ranking," *J. Acoust. Soc. Amer.*, vol. 132, no. 1, pp. 392–402, Jul. 2012.

[19] 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.

[20] 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.

[21] M. J. Lindenbeck, B. Laback, P. Majdak, and S. Srinivasan, "Temporal-pitch sensitivity in electric hearing with amplitude modulation and inserted pulses with short inter-pulse intervals," *J. Acoust. Soc. Amer.*, vol. 147, no. 2, pp. 777–793, Feb. 2020.

[22] R. J. van Hoesel, "A peak-derived timing stimulation strategy for a multichannel cochlear implant," Patent PCT/AU2002/000660, 2002.

[23] C. M. Zierhofer, "Electrical nerve stimulation based on channel specific sampling sequences," U.S. Patent 6594525, Jul. 15, 2003.

[24] I. Hochmair *et al.*, "MED-EL cochlear implants: State of the art and a glimpse into the future," *Trends Amplif.*, vol. 10, no. 4, pp. 201–219, Dec. 2006.

[25] T. Fischer, C. Schmid, M. Kompis, G. Mantokoudis, M. Caversaccio, and W. Wimmer, "Effects of temporal fine structure preservation on spatial hearing in bilateral cochlear implant users," *J. Acoust. Soc. Amer.*, vol. 150, no. 2, pp. 673–686, Aug. 2021.

[26] V. Müller, H. Klünter, D. Fürstenberg, H. Meister, M. Walger, and R. Lang-Roth, "Examination of prosody and timbre perception in adults with cochlear implants comparing different fine structure coding strategies," *Amer. J. Audiol.*, vol. 27, no. 2, pp. 197–207, Jun. 2018.

[27] E. H.-H. Huang, C.-M. Wu, and H.-C. Lin, “Combination and comparison of sound coding strategies using cochlear implant simulation with Mandarin speech,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 2407–2416, 2021.

[28] W. Huang, L. L. N. Wong, and F. Chen, “Just-noticeable differences of fundamental frequency change in Mandarin-speaking children with cochlear implants,” *Brain Sci.*, vol. 12, no. 4, p. 443, Mar. 2022.

[29] Q. Meng, N. Zheng, and X. Li, “A temporal limits encoder for cochlear implants,” in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2015, pp. 5863–5867.

[30] Q. Meng, N. Zheng, and X. Li, “Mandarin speech-in-noise and tone recognition using vocoder simulations of the temporal limits encoder for cochlear implants,” *J. Acoust. Soc. Amer.*, vol. 139, no. 1, pp. 301–310, Jan. 2016.

[31] Q. Meng, X. Wang, N. Zheng, J. W. H. Schnupp, and A. Kan, “Binaural hearing measured with the temporal limits encoder using a vocoder simulation of cochlear implants,” *Acoust. Sci. Technol.*, vol. 41, no. 1, pp. 209–213, 2020.

[32] 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.

[33] A. Kan and Q. Meng, “The temporal limits encoder as a sound coding strategy for bilateral cochlear implants,” *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 29, pp. 265–273, 2021.

[34] R. Ghosh, H. Ali, and J. H. L. Hansen, “CCi-MOBILE: A portable real time speech processing platform for cochlear implant and hearing research,” *IEEE Trans. Biomed. Eng.*, vol. 69, no. 3, pp. 1251–1263, Mar. 2022.

[35] R. C. M. C. Shekar and J. H. L. Hansen, “An evaluation framework for research platforms to advance cochlear implant/hearing aid technology: A case study with CCi-MOBILE,” *J. Acoust. Soc. Amer.*, vol. 149, no. 1, pp. 229–245, Jan. 2021.

[36] J.-N. Li *et al.*, “The advances in hearing rehabilitation and cochlear implants in China,” *Ear Hearing*, vol. 38, no. 6, p. 647, 2017.

[37] H. Levitt, “Transformed up-down methods in psychoacoustics,” *J. Acoust. Soc. Amer.*, vol. 49, no. 2B, pp. 467–477, 1971.

[38] C. Micheyl, K. Delhommeau, X. Perrot, and A. J. Oxenham, “Influence of musical and psychoacoustical training on pitch discrimination,” *Hearing Res.*, vol. 219, nos. 1–2, pp. 36–47, Sep. 2006.

[39] B. R. Glasberg and B. C. Moore, “Psychoacoustic abilities of subjects with unilateral and bilateral cochlear hearing impairments and their relationship to the ability to understand speech,” *Scand. Audiol.*, vol. 32, pp. 1–25, Jan. 1988.

[40] M. K. Qin and A. J. Oxenham, “Effects of envelope-vocoder processing on F0 discrimination and concurrent-vowel identification,” *Ear Hearing*, vol. 26, no. 5, pp. 451–460, 2005.

[41] V. Summers and M. R. Leek, “F0 processing and the seperation of competing speech signals by listeners with normal hearing and with hearing loss,” *J. Speech, Lang., Hearing Res.*, vol. 41, no. 6, pp. 1294–1306, Dec. 1998.

[42] R. L. Goldsworthy, “Correlations between pitch and phoneme perception in cochlear implant users and their normal hearing peers,” *J. Assoc. Res. Otolaryngol.*, vol. 16, no. 6, pp. 797–809, Dec. 2015.

[43] 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.

[44] R. L. Goldsworthy, L. A. Delhorne, L. D. Braida, and C. M. Reed, “Psychoacoustic and phoneme identification measures in cochlear-implant and normal-hearing listeners,” *Trends Amplification*, vol. 17, no. 1, pp. 27–44, Mar. 2013.

[45] M. Marx *et al.*, “Speech prosody perception in cochlear implant users with and without residual hearing,” *Ear Hearing*, vol. 36, no. 2, pp. 239–248, 2015.

[46] V. D. Tejani and C. J. Brown, “Speech masking release in hybrid cochlear implant users: Roles of spectral and temporal cues in electric-acoustic hearing,” *J. Acoust. Soc. Amer.*, vol. 147, no. 5, pp. 3667–3683, May 2020.

[47] J. C. Kopelovich, M. D. Eisen, and K. H. Franck, “Frequency and electrode discrimination in children with cochlear implants,” *Hearing Res.*, vol. 268, nos. 1–2, pp. 105–113, Sep. 2010.

[48] N. T. Jiam, M. L. Deroche, P. Jiradejvong, and C. J. Limb, “A randomized controlled crossover study of the impact of online music training on pitch and timbre perception in cochlear implant users,” *J. Assoc. Res. Otolaryngol.*, vol. 20, no. 3, pp. 247–262, Jun. 2019.

[49] R. Kang *et al.*, “Development and validation of the university of Washington clinical assessment of music perception test,” *Ear Hearing*, vol. 30, no. 4, p. 411, 2009.

[50] A. Vandali, D. Sly, R. Cowan, and R. Van Hoesel, “Training of cochlear implant users to improve pitch perception in the presence of competing place cues,” *Ear Hearing*, vol. 36, no. 2, pp. e1–e13, 2015.

[51] A. Krenmayr, D. Visser, R. Schatzer, and C. Zierhofer, “The effects of fine structure stimulation on pitch perception with cochlear implants,” *Cochlear Implants Int.*, vol. 12, no. 1, pp. S70–S72, May 2011.

[52] C. Zierhofer and R. Schatzer, “A fine structure stimulation strategy and related concepts,” in *Cochlear Implant Research Updates*. Norderstedt, Germany, 2012, p. 91.

[53] M. F. Dorman and P. C. Loizou, “Mechanisms of vowel recognition for ineraid patients fit with continuous interleaved sampling processors,” *J. Acoust. Soc. Amer.*, vol. 102, no. 1, pp. 581–587, Jul. 1997.

[54] K. Vermeire, A. K. Punte, and P. Van de Heyning, “Better speech recognition in noise with the fine structure processing coding strategy,” *ORL*, vol. 72, no. 6, pp. 305–311, 2010.

[55] J. Laneau, J. Wouters, and M. Moonen, “Relative contributions of temporal and place pitch cues to fundamental frequency discrimination in cochlear implantees,” *J. Acoust. Soc. Amer.*, vol. 116, no. 6, pp. 3606–3619, Dec. 2004.

[56] X. Gao, D. B. Grayden, and M. D. McDonnell, “Modeling electrode place discrimination in cochlear implant stimulation,” *IEEE Trans. Biomed. Eng.*, vol. 64, no. 9, pp. 2219–2229, Sep. 2017.