Binaural hearing measured with the temporal limits encoder using a vocoder simulation of cochlear implants

Qinglin Meng1,2,*, Xianren Wang3, Nengheng Zheng4, Jan W. H. Schnupp2 and Alan Kan5,6

1Acoustics Lab of School of Physics and Optoelectronics and State Key Laboratory of Subtropical Building Science, South China University of Technology, China
2Hearing Research Group, Department of Biomedical Sciences, City University of Hong Kong, Hong Kong SAR of China
3Department of Otorhinolaryngology, the First Affiliated Hospital, Sun Yat-Sen University and Institute of Otorhinolaryngology, Sun Yat-Sen University, China
4The Guangdong Key Laboratory of Intelligent Information Processing, College of Information Engineering, Shenzhen University, China
5Waisman Center, University of Wisconsin-Madison, WI, USA

Abstract: Cochlear implants (CIs) convert sound to electrical stimulation by extracting the envelope in each frequency band while discarding the temporal fine structure (TFS). This processing removes the fine structure interaural time differences (ITDs), which are an important cue for locating sounds on the horizontal plane in normal-hearing (NH) listeners, but are unavailable to CI users. A temporal limits encoder (TLE) strategy was previously proposed to enhance TFS in CIs, and our previous studies via tone-carrier vocoder simulation have shown improved unilateral speech-in-noise understanding and pitch perception. Here, binaural benefits of TLE were assessed, by measuring the binaural intelligibility level difference (BILD), using a 22-channel tone-carrier vocoder in NH listeners. TLE was compared to continuous interleaved sampling (CIS), a common CI strategy. Speech reception thresholds (SRTs) were measured for diotic target speech (male), and diotically-colocated or dichotically-separated (applying 625 ms delay between ears) competitors (male or female).

Compared to CIS, TLE showed significantly larger BILDs for different genders, indicating that TLE-simulation listeners were able to benefit from both pitch and spatial cues. However, SRTs for vocoded conditions were much higher than non-vocoded listening, likely due to a lack of familiarity with vocoded speech listening.

Keywords: Binaural hearing, Cochlear implant, Vocoder, Temporal fine structure

PACS number: 43.64.Me, 43.66.Pn, 43.66.Ts [doi:10.1250/ast.41.209]

1. INTRODUCTION

People usually listen with two ears, but until recently, hundreds of thousands of patients with severe-to-profound hearing-loss have been implanted with an electronic device, i.e., cochlear implant (CI), in one ear only to regain some sound perception abilities. Bilateral cochlear implant (BiCI) user numbers have been fairly small up to the early 2000s [1], but are becoming increasingly common [2]. Compared with unilateral CI listening, BiCI users generally show improved performance in localization and speech-in-noise recognition. However, their performance is still significantly poorer than normal hearing (NH) listeners [3,4]. One important reason is that BiCIs cannot provide users access to normal binaural cues, especially fine structure interaural time differences (ITDs). Most current CIs modulate a constant high pulse rate (~1,000 pulse-per-second,pps) with a signal envelope to convey sound to the listener, but without precise interaural timing synchronization and frequency (or place) -matching [5–7]. Therefore, for each frequency channel, the ITDs can only be weakly represented in the envelope but not in the temporal fine structure (TFS), which NH listeners rely on for sound localization judgements [8] especially in noisy environments [9,10].

To enhance ITDs with electrical stimulation, most strategies lower the stimulation rate of the low frequency channels by timing the firing of the pulses at the time of zero-crossing [11] or peak [7] of the corresponding bands’
signal, but only small binaural benefits have been observed [12]. Recently, some proof-of-concept studies have proposed novel approaches to improving ITD sensitivity in BiCl users: combining beamformer-based direction estimation with peak timing [13]; introducing some pulses with short interpulse intervals in high stimulation rates to promote ITD sensitivity [14]; interleaving high and low rate stimulation on different channels [15]. All studies showed significant improvement when measuring ITD sensitivity, but there are two main limitations: 1. no speech intelligibility tests were carried out; and 2. their applications might need a real-time ITD estimation algorithm which is a difficult engineering task for real scenarios.

In the current study, we tested a novel BiCl strategy based on the temporal limits encoder (TLE) strategy [16,17]. The idea of TLE is to enhance the TFS representation by downward frequency transposition of higher-frequency channel signals to a lower frequency. The transposed signal maintains the overall temporal envelope but also introduces a slowly varying TFS whose fluctuation rate is within a range defined by the psychophysical temporal pitch limits. This technique is unique in that it does not require an explicit envelope/TFS decomposition step. Previous vocoder simulation experiments have shown some advantages in pitch and speech-in-noise tasks when listening with TLE unilaterally [16,17]. We hypothesized that the slow-varying TFS introduced by a bilateral TLE (Bi-TLE) strategy may enhance ITD cues, and thereby promote binaural benefit in speech-in-noise understanding. To test this hypothesis, we conducted a binaural intelligibility level difference (BILD) experiment, comparing TLE to the continuous interleaved sampling (CIS) strategy (a common strategy used in Cls). A vocoder was used to simulate Cl listening for NH listeners in this experiment. As a proof-of-concept, NH listeners were tested to avoid the inherent performance variability typically found in Cl listeners.

2. BILATERAL TEMPORAL LIMITS ENCODER ALGORITHM

Here, we describe briefly the TLE strategy in a binaural hearing scenario (more details about unilateral TLE can be found in [16]). We assume that the incoming sound at the left ear arrives earlier than the right ear by a difference of ITD. A band-pass filter bank is used to analyze the sound signal and extract a modulator for the electrical pulse train. Figure 1 illustrates the modulator extraction processes where $x_k(t)$ denotes the band signal from the $k$th band-pass filter, and $m_k(t)$ denotes the derived modulator for the $k$th electrodes.

From the equations in the figure, we can see that Bi-CIS only preserves ITD between the envelopes of the modulators from both sides, while Bi-TLE is able to preserve the ITD in the envelopes, as well as the ITD in the derived modulator TFS in the two ears. This is achieved by a frequency downshifting technique which generates a new, slowly-varying, band-limited modulator which implicitly preserves the TFS in each side, and the interaural phase difference (IPD) between the two ears. Hence, no TFS or ITD estimation is used in the Bi-TLE strategy. The Bi-TLE derived modulators is then used to amplitude modulate the electrical pulses in a Cl in each ear, respectively.

![Fig. 1](image)

*Fig. 1* The signal processing flow chart for the amplitude modulator extraction in Bi-CIS and Bi-TLE. $x_k(t)$: sound signal from $k$th band-pass filter; $m_k(t)$: pulse train amplitude modulator; $e_k(t)$: temporal envelope; $\cos()$: temporal fine structure; ear side: L and R; The sound is assumed to first arrive at left ear and then at right ear and the delay is ITD.
3. Vocoder Simulation in NH

3.1. Subjects and Material

Seven NH students were recruited from Sun Yat-Sen University. Participants were compensated and all subjects gave informed consent in accordance with the local institution’s review board. We used sentence material from the Mandarin Hearing-In-Noise Test (MHINT) corpus [18], which comprises of 12 test lists and 2 training lists. Each list consisted of twenty 10-mono-syllabic-word sentences spoken by a male speaker. The sentences were presented at a 16 kHz sampling rate.

3.2. Vocoder Implementation

For each ear, sixth-order Butterworth band-pass filters were used to divide the speech signal into 22 channels within the frequency range of 80 to 7,999 Hz. The cutoff frequencies of the filters were chosen according to the Greenwood function [19]. They were 80.0, 122.4, 172.1, 230.4, 298.7, 378.8, 472.8, 583.1, 712.3, 863.9, 1,041.6, 1,250.1, 1,494.5, 1,781.2, 2,117.3, 2,511.5, 2,973.8, 3,515.9, 4,151.7, 4,897.2, 5,771.5, 6,796.7, and 7,999.0 Hz.

For the CIS strategy, a half-wave rectification followed by an eighth-order Butterworth low-pass filter with a cutoff frequency of 125 Hz was used to extract the temporal envelope as shown in Fig. 1. For the TLE strategy, the within-channel processing method was dependent on the bandwidth and higher-edge cutoff frequency. For the three lowest-frequency channels, the band passed signals were directly used as the modulator for pulse modulation. For the middle frequency channels (channels 4 to 12; covering a frequency range from 230.4 to 1,494.5 Hz), the TLE processing method shown in Fig. 1 was used; the down-modulator frequency, \( f_m \), was set to the lower cutoff frequency of the corresponding channel minus 50 Hz [16]; the low-pass filter was an eighth-order Butterworth filter with the cutoff frequency set at the bandwidth of corresponding channel + 150 Hz. For the remaining channels (13–22), CIS envelope extraction was used. For both strategies, each modulator, \( m_k \), was used to amplitude-modulate a sinusoidal signal whose frequency was at the center of the corresponding channel, and its initial phase was fixed at zero. Finally, the modulated signals from all bands were summed together to generate vocoded speech.

3.3. Binaural Intelligibility Level Difference (BILD) Experiment

A BILD test was used in this experiment as a proof-of-concept to understand whether the ITD cues introduced by the Bi-TLE strategy would lead to a binaural benefit. A BILD experiment is useful here because it avoids the influence of other cues that might contribute to a binaural benefit, such as interaural level differences, head shadow, and binaural squelch. In this experiment, an ITD was introduced into the acoustic signal prior to vocoder processing to simulate real-world listening. The test stimulus consisted of a diotic male target, and four competitor configurations: \( M_0 \): diotic two male competitors, both with an ITD of 0 \( \mu \)s; \( M_{625} \): one male competitor with a left-leading ITD of 625 \( \mu \)s, and the other male competitor with a right-leading ITD of 625 \( \mu \)s; \( F_0 \): diotic two female competitors, both with an ITD of 0 \( \mu \)s; and \( F_{625} \): one female competitor with a left-leading ITD of 625 \( \mu \)s and the other female competitor with a right-leading ITD of 625 \( \mu \)s. An ITD of 625 \( \mu \)s was chosen because it corresponds to a 10-sample delay at 16 kHz sampling frequency and is close to the physiological maximum of an adult head. The competitors were generated by concatenating ten sentences from MHINT Practice List 2, which is spoken by a male speaker. The concatenated male speech was transformed to female-like speech using the STRAIGHT toolbox [20] by increasing the pitch by 100 Hz. For each trial, two random section of the concatenated competitor speech was chosen. The competitor always started playing 300 ms before the target talker and end 300 ms after. In total, 12 conditions were tested (3 listening conditions (non-vocoded, Bi-CIS vocoded, Bi-TLE vocoded \( \times \) 4 competitor configurations) for each subject. The condition order and list order were both randomized for each subject.

The speech reception threshold (SRT) was measured for each condition using a 1-down-1-up adaptive procedure. The initial SNR was 10 dB. The SNR was changed by 8 dB before the second reversal, 4 dB before the fourth reversal, and then by 2 dB until the end of the test list. The SNR among the last 8 sentences were averaged to estimate the SRT [16]. For each trial, the subject was instructed to repeat as many words as possible. When the subject was not able to repeat every word in the sentence correctly, they were given one further presentation and attempt with the same sentence. A sentence was scored as correct when eight or more words were recalled correctly. For each strategy, the BILD was calculated for each gender competitor by taking the difference between SRT values for the diotic and dichotic competitor conditions, e.g., \( SRT_{M_0} - SRT_{M_{625}} \). An informal practice session (using about 10 sentences, which were not used in the formal session) was carried out for each subject to get used to the vocoded clear speech.

3.4. Results

The SRTs are shown in Fig. 2 for the 12 conditions. In all conditions, SRTs for the non-vocoded conditions were lower than the vocoded conditions. In addition, SRTs for configurations with female competitors were usually lower.
than when there were male competitors. These observations are somewhat to be expected because listeners had less familiarity with listening to vocoded speech. In the case of a female competitor, lower SRTs may be due to an additional pitch cue difference between the target and competitors which makes the task easier. Alternatively, a less likely explanation is that the use of the STRAIGHT toolbox for creating the female speech may have introduced a distortion that made the competitor less intelligible, thereby reducing the effects of informational masking which in turn would lead to better performance. Because target-competitor gender differences were expected to affect outcomes, separate repeated-measures analysis of variance (ANOVA) were conducted for data collected for each gender competitor to compare the effect of vocoder condition on SRTs for different competitor locations. For both gender competitors, significant main effects were found for vocoder condition [male: $F(2) = 63.13$, $p < 0.001$; female: $F(2) = 210.57$, $p < 0.001$] and competitor location [male: $F(1) = 74.97$, $p < 0.001$; female: $F(1) = 15.89$, $p = 0.007$]. However, there was a significant interaction between vocoder condition and competitor location [male: $F(2) = 27.3$, $p < 0.001$; female: $F(2) = 10.55$, $p = 0.002$] suggesting that the main effect of competitor location is dependent on vocoder condition. Pairwise comparison of competitor location SRTs for each vocoder condition revealed that a significant difference exists for competitor location for the non-vocoded and Bi-TLE conditions ($p < 0.001$) but not Bi-CIS for a male competitor, while a significant difference for competitor location was only found in the Bi-TLE condition for a female competitor ($p < 0.001$). In addition, post-hoc analysis with Bonferroni correction revealed that SRTs for the non-vocoded condition was significantly better than the two vocoded conditions, while no significant difference was found between Bi-TLE and Bi-CIS.

We were particularly interested in understanding whether TFS-ITD cues encoded by Bi-TLE provided any benefit over envelope ITD cues encoded by Bi-CIS. Average BILDs were 8.4, −0.6 and 4.1 dB for non-vocoded, Bi-CIS and Bi-TLE, respectively when the competitor was male, and 1.6, 0.1, and 4.2 dB when the competitor was female. Repeated-measures ANOVA revealed a statistically significant difference in the BILDs for vocoder condition [male: $F(2) = 27.3$, $p < 0.001$; female: $F(2) = 10.55$, $p = 0.002$]. Post-hoc analysis revealed that for a male competitor, BILDs for non-vocoded, Bi-CIS and Bi-TLE were all significantly different from each other, while for a female competitor, BILDs for Bi-TLE were significantly different from the non-vocoded and Bi-CIS conditions. These results suggest that Bi-TLE provides a relatively consistent BILD benefit irrespective of the gender of the competitor, though this benefit is still smaller than that obtained by non-vocoded, normal hearing.

3.5. Discussion

TLE transposes original fast TFS within each channel to a slowly varying version, without explicit extraction or encoding of ITDs. In this experiment, we found that bilateral implementation of TLE can provide ~4 dB of binaural benefit. In contrast, no binaural benefit was found with CIS. However, the benefit with TLE is still smaller than that of non-vocoded normal hearing, suggesting that there is still a difference between the TFS ITD introduced by Bi-TLE and that found in the acoustic signal. This difference may be due to the choice of $f_z$ and the low-pass filter cutoff used in this experiment, and further work is needed to determine the best choice for these parameters.

There are a number of caveats to note for this experiment. While a binaural benefit was observed for Bi-TLE, the testing conditions used in this experiment is not the same as real world listening for bilateral CI users. In this experiment, we only provided ITD cues for spatial separation and in the real world, there is an interaction of ITD and ILD cues which may change the outcome of this experiment. Further, differences in insertion depths of electrode arrays in the two ears has been shown to affect sensitivity to ITDs [21,22], as well as binaural benefits such as spatial release from masking [23]. Hence, in bilateral CI users, the benefits of Bi-TLE may be smaller than that observed in this study with NH listeners. In addition, sampling clocks in the CI processors at the two ears are usually not synchronized, and the effect of sampling clock

![Graph](image-url)
differences were not simulated. These factors should be examined in future testing to understand the robustness of Bi-TLE for providing binaural benefit in speech-in-noise listening.

4. CONCLUSION

Significant binaural benefits were found for TLE but not for CIS in a BILD experiment with a 22-channel sine-wave vocoder. The results suggest that Bi-TLE may be a promising bilateral CI strategy for improving speech-in-noise understanding by providing a binaural benefit. We hope to evaluate this in future testing with bilateral CI listeners.

ACKNOWLEDGEMENTS

We thank all the subjects who participated in these experiments. This work is jointly supported by NSF of China [Grant No. 11704129 and 61771320], the Fundamental Research Funds for the Central Universities (SCUT), State Key Laboratory of Subtropical Building Science [SCUT, Grant No. 2018ZB23], and Shenzhen Science and Innovation Funds [Grant No. JCYJ 20170302145906843].

REFERENCES

[1] R. van Hoesel, R. Ramsden and M. O’Driscoll, “Sound-direction identification, interaural time delay discrimination, and speech intelligibility advantages in noise for a bilateral cochlear implant user,” *Ear Hear.*, 23, 137–149 (2002).

[2] B. R. Peters, J. Wyss and M. Manrique, “Worldwide trends in bilateral cochlear implantation,” *Laryngoscope*, 120, Suppl:S17-44, doi:10.1002/lary.20859 (2010).

[3] S. Kerber and B. U. Seeber, “Sound localization in noise by normal-hearing listeners and cochlear implant users,” *Ear Hear.*, 33, 445–457 (2012).

[4] P. C. Loizou, Y. Hu, R. Y. Litovsky, G. Yu, R. Peters, J. Lake and P. Roland, “Speech recognition by bilateral cochlear implant users in a cocktail-party setting,” *J. Acoust. Soc. Am.*, 125, 372–383 (2009). doi:10.1121/1.3036175

[5] A. Kan and R. Y. Litovsky, “Binaural hearing with electrical stimulation,” *Hear. Res.*, 322, 127–137 (2015).

[6] B. Laback, K. Egger and P. Majdak, “Perception and coding of interaural time differences with bilateral cochlear implants,” *Hear. Res.*, 322, 138–150 (2015).

[7] R. van Hoesel, “Bilateral Cochlear Implants,” in *Auditory Prostheses: New Horizons*, F. Zeng, A. N. Popper and R. R. Fay, Eds. (Springer, New York, 2012), pp. 13–57.

[8] E. A. Macpherson and J. C. Middlebrooks, “Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited,” *J. Acoust. Soc. Am.*, 111, 2219–2236, doi:10.1121/1.1471898 (2002).

[9] C. Zhang, L. Lu, X. Wu and L. Li, “Attentional modulation of the early cortical representation of speech signals in informational or energetic masking,” *Brain Lang.*, 135, 85–95, doi:10.1016/j.bandl.2014.06.002 (2014).

[10] C. Zhang, R. Tao and H. Zhao, “Auditory spatial attention modulates the unmasking effect of perceptual separation in a “cocktail party” environment,” *Neuropsychologia*, 124, 108–116, doi:10.1016/j.neuropsychologia.2019.01.009 (2019).

[11] I. Hochmair, P. Nopp, C. Jolly, M. Schmidt, H. Schöller, C. Garnham and I. Anderson, “MED-EL cochlear implants: State of the art and a glimpse into the future,” *Trends Amplif.*, 10, 201–219 (2006).

[12] S. Zirn, S. Arndt, A. Aschendorff, R. Laszig and T. Wesarg, “Perception of interaural phase differences with envelope and fine structure coding strategies in bilateral cochlear implant users,” *Trends Hear.*, 20, 1–12 (2016).

[13] B. Williges, T. Jurgens, H. Hu and M. Dietz, “Coherent coding of enhanced interaural cues improves sound localization in noise with bilateral cochlear implants,” *Trends Hear.*, 22, 1–18 (2018).

[14] S. Srinivasan, B. Laback, P. Majdak and B. Delgutte, “Introducing short interpulse intervals in high-rate pulse trains enhances binaural timing sensitivity in electric hearing,” *J. Assoc. Res. Otalaryngol.*, 19, 301–315 (2018).

[15] T. Thakkar, A. Kan, H. G. Jones and R. Y. Litovsky, “Mixed stimulation rates to improve sensitivity of interaural timing differences in bilateral cochlear implant listeners,” *J. Acoust. Soc. Am.*, 143, 1428–1440 (2018).

[16] 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. Am.*, 139, 301–310 (2016).

[17] Q. Meng, N. Zheng and X. Li, “A temporal limits encoder for cochlear implants,” *Proc. ICASSP 2015*, pp. 5863–5867 (2015).

[18] L. L. Wong, S. D. Soli, S. Liu, N. Han and M. W. Huang, “Development of the Mandarin Hearing in Noise Test (MHINT),” *Ear Hear.*, 28, 70S–74S (2007).

[19] D. D. Greenwood, “A cochlear frequency-position function for several species—29 years later,” *J. Acoust. Soc. Am.*, 87, 2592–2605 (1990).

[20] H. Kawahara, H. Banno, T. Irimo and P. Zolfaghari, “Algorithm amalgam: Morphing waveform based methods, sinusoidal models and STRAIGHT,” *Proc. ICASSP 2004*, pp. 1–13 (2004).

[21] A. Kan, R. Y. Litovsky and M. J. Goupell, “Effects of interaural pitch-matching and auditory image centering on binaural sensitivity in cochlear-implant users,” *Ear Hear.*, 36, e62–e68 (2015).

[22] A. Kan, C. Stoelb, R. Y. Litovsky and M. J. Goupell, “Effect of mismatched place-of-stimulation on binaural fusion and lateralization in bilateral cochlear-implant users,” *J. Acoust. Soc. Am.*, 134, 2923–2936 (2015).

[23] M. J. Goupell, C. A. Stoelb, A. Kan and R. Y. Litovsky, “The effect of simulated interaural frequency mismatch on speech understanding and spatial release from masking,” *Ear Hear.*, 39, 895–905, doi:10.1097/AUD.0000000000000541 (2018).