



Enhancing the Interaural Time Difference of Bilateral Cochlear Implants with the Temporal Limits Encoder

Yangyang Wan¹, Huali Zhou², Qinglin Meng², Nengheng Zheng^{1,3}

¹The Guangdong Key Laboratory of Intelligent Information Processing, College of Electronics and Information Engineering, Shenzhen University, China

²Acoustics Laboratory, School of Physics and Optoelectronics, South China University of Technology, China

³Peng Cheng Laboratory, Shenzhen, China

2170269139@email.szu.edu.cn, mshualizhou@mail.scut.edu.cn,
mengqinglin@scut.edu.cn, nhzheng@szu.edu.cn

Abstract

Normal hearing listeners mainly use interaural time differences (ITDs) and interaural level differences (ILDs) to localize sound sources in the horizontal plane. Listeners with bilateral cochlear implants (CIs), however, have poor sensitivity to ITDs which significantly limits their spatial hearing capabilities. Most CI signal processing strategies, such as the continuous interleaved sampling (CIS) strategy, are temporal-envelope-based, and the temporal fine structure (TFS), which contains useful cues for ITDs, is discarded. Recently, a temporal limits encoder (TLE) CI strategy was proposed to implicitly introduce the TFS while preserving the temporal envelope. It has demonstrated benefits in unilateral CI simulations in tasks including speech-in-noise understanding and pitch perception. This study investigates whether the ITD cues could be enhanced by a bilateral TLE strategy. Identification of five ITDs respectively associated with five sound source directions was tested with vocoded speech stimuli to compare the performance of the bilateral TLE and CIS strategies. Results show that the bilateral TLE has better overall performance than the bilateral CIS. This finding suggests that the bilateral TLE is promising in providing enhanced ITD cues for bilateral CI users.

Index Terms: sound source localization, interaural time difference, bilateral cochlear implants, temporal fine structure

1. Introduction

Cochlear implants (CIs) have helped hundreds of thousands of people with severe-to-profound hearing loss to regain hearing [1]. Although unilateral CIs can provide most users good speech perception under favorable conditions, bilateral cochlear implantations are becoming more and more common with the increasing expectation of binaural benefits in speech perception in noise and spatial hearing [2, 3].

Normal-hearing (NH) listeners rely on binaural cues, mainly the interaural timing difference (ITD) and interaural level difference (ILD), to localize sound sources in the horizontal plane and to segregate a target speech from interfering sounds in the environment. Most modern CI strategies extract only the temporal envelopes, whereas the temporal fine structure (TFS) is discarded. Between TFS of the two ears, there are important ITD cues (i.e., TFS-ITD) for NH listeners to judge the sound locations. For bilateral CIs, the ITD cues are only conveyed in the envelopes between the two sides

(i.e., envelope-ITD). The lack of bilateral pulse synchronization and the frequency-to-electrode mismatch are also limiting factors. Therefore, Bi-CI users cannot make full use of the ITD cues to localize sounds and consequently have abnormal spatial hearing abilities [4, 5].

Many efforts have been made to enhance the ITD cues. One approach is to explicitly encode the TFS at the low-frequency channels by controlling the pulse timing, but limited binaural benefits were observed [6]. In recent years, some proof-of-concept strategies have been proposed, e.g. introducing short inter-pulse interval in high-rate pulse trains [7] and using mixed stimulation rates [8]. Although ITD sensitivity might be improved, the integration of these algorithms needs a real-time ITD estimation, which is a challenging engineering task especially when multiple sounds are received at the same time.

In this study, we tested a new bilateral CI strategy, namely the bilateral temporal limits encoder (Bi-TLE). For unilateral CIs, TLE aims at improving the TFS representation by a frequency transposition of high-frequency bandlimited signals to a low-frequency range, which falls in the range of temporal pitch limits with electric hearing (~50-300 Hz for most CI users) [9, 10]. The TLE's frequency transposition, replacing the envelope extraction stage of a standard continuous interleaved sampling (CIS) strategy, generates a novel amplitude modulator which preserves the temporal envelope and introduces a new slowly varying TFS in an implicit manner. This technique is unique in that it does not require an explicit envelope/TFS decomposition. In previous vocoder simulation experiments, some improvements have been observed in a speech-in-noise recognition and pitch perception tasks when listening with unilateral TLE strategy [9, 10]. In a spatial-release-from-masking test with actual bilateral CI users [11] and a binaural intelligibility level difference test with simulated bilateral CI users [12], Bi-TLE also showed better performance across the cohorts than standard envelope-based strategies. These results imply that the ITD cues may be enhanced by TLE. In Bi-TLE, ITDs are not explicitly extracted, but the introduced slowly varying TFS may implicitly contain ITD cues between ears. In this study, we designed a sound localization experiment to compare the performance of Bi-TLE and Bi-CIS. Five ITDs associated with five azimuths (0° , $\pm 40^\circ$, and $\pm 80^\circ$) respectively were used as the delays between the two ears. The localization task can provide more direct evidence on the feasibility of ITDs in spatial hearing for bilateral CI users. If Bi-TLE shows better performance than bilateral CIS (Bi-CIS) in

the experiment, the results will further suggest that TFS-ITDs will likely be able to be enhanced by the Bi-TLE strategy.

2. Algorithm

2.1. TLE and CIS algorithms

Figure 1 shows the signal processing flow charts of the amplitude modulator generation for the k th channel in CIS and TLE strategies, where $x_k(t)$ represents the sub-band signal of the k th channel after the original sound signal is filtered through the k th band-pass filter. In Figure 1(a), the bandwidth-limited signal $x_k(t)$ is processed by a rectifier and a low-pass filter to obtain the k th channel's temporal envelope, $e_k(t)$, which will be taken as the CIS modulator. In Figure 1(b), where B_k denotes the effective bandwidth of the sub-band signal $x_k(t)$ and C_k represents the bandwidth of the temporal pitch limits range (e.g., 250 Hz). For frequency bands with a bandwidth narrower than the temporal pitch limits range (i.e., $B_k \leq C_k$), $x_k(t)$ is multiplied by a sinusoidal signal $\cos(2\pi f_k t)$ and then passed through a low-pass filter to generate a new signal $u_k(t)$, which is a frequency downshifted version of $x_k(t)$ and used as the TLE modulator. The sub-band signals of the wide-band channels (i.e., $B_k > C_k$) are processed in the way as in the CIS strategy to generate the temporal envelope signals $e_k(t)$.

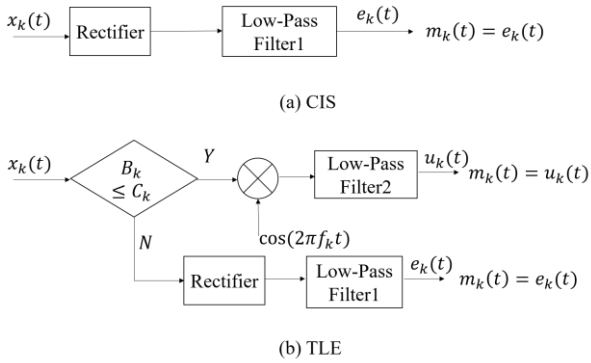


Figure 1: Signal processing flow charts of the amplitude modulator generation for the k th channel in CIS(a) and TLE(b) strategy.

The lower cutoff frequency of the k th channel' sub-band signal is f_{kl} , which is defined by the frequency allocation of CI fitting. The frequency of the lower limit of temporal pitch perception on the k th electrode of the CI is F_L , which is about 50 Hz. The frequency of the sinusoidal $\cos(2\pi f_k t)$, namely f_k , is set at $f_{kl} - F_L$. It ensures that the frequency components of the $u_k(t)$ is above F_L . More details about the TLE strategy can be found in [9] and [10].

According to signal processing theory [13, 14], any sub-band signal can be represented in the form of a quasi-sinusoidal oscillation, i.e.,

$$x_k(t) = e_k(t) \cdot \cos(2\pi f_{ck} t + \varphi_k(t)) \quad (1)$$

where $e_k(t)$ is the temporal envelope and $2\pi f_{ck} t + \varphi_k(t)$ is the phase. Let $y_k(t)$ be a sinusoidal carrier, i.e., $y_k(t) = \cos(2\pi f_k t)$, it is quite straightforward to see that

$$x_k(t) \cdot y_k(t) = e_k(t) \cdot \cos(2\pi f_{ck} t + \varphi_k(t)) \cdot \cos(2\pi f_k t)$$

$$= \frac{e_k(t)}{2} \cdot \cos(2\pi(f_{ck} - f_k)t + \varphi_k(t)) + \frac{e_k(t)}{2} \cdot \cos(2\pi(f_{ck} + f_k)t + \varphi_k(t)) \quad (2)$$

Applying a low-pass filter to the signal given by (2), we get

$$u_k(t) = \frac{e_k(t)}{2} \cdot \cos(2\pi(f_{ck} - f_k)t + \varphi_k(t)) \quad (3)$$

In this way, the envelope $e_k(t)$ is well preserved and the sub-band signal $x_k(t)$ around the high frequency of f_{ck} is downshifted to a low frequency of “ $f_{ck} - (f_{kl} - F_L)$ ”.

2.2. Using TLE at both ear sides (Bi-TLE)

Here, we chose a complex signal consisting of three components whose frequencies are 540, 550 and 560 Hz, to show the difference in CIS and TLE strategy processing (see Figure 2). We can see that the original signal contains the Hilbert envelope and temporal fine structure. As shown in the bottom panel, CIS only extracts the temporal envelope of the original signal, and the TFS information is discarded. After the TLE strategy, however, not only is the Hilbert envelope of the original signal retained but also a new slowly varying TFS is introduced, as illustrated in the middle panel.

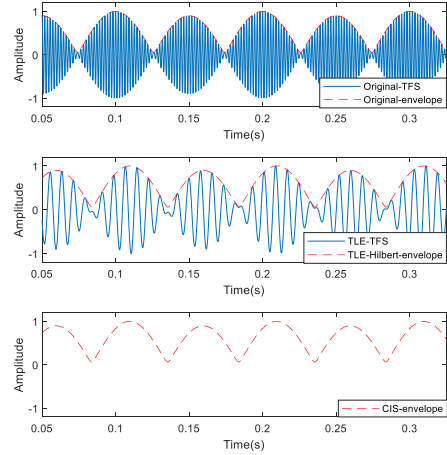


Figure 2: The waveforms of a complex signal processed by CIS and TLE strategy

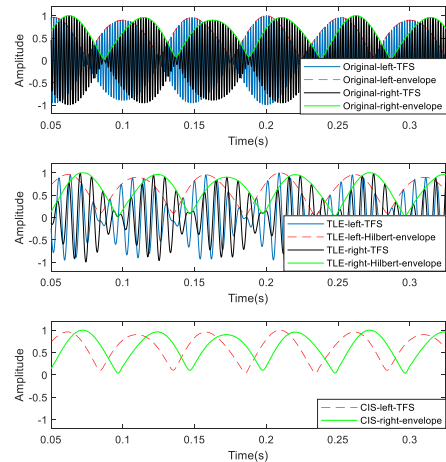


Figure 3: The waveforms of a binaural signal with an ITD of 625 μ s processed by Bilateral TLE and CIS

Figure 3 illustrates how a binaural signal with an ITD of $625 \mu\text{s}$ is processed by the bilateral CIS and TLE strategies. We can see that the ITD is only conveyed between the temporal envelopes of the left and right ears with Bi-CIS. In Bi-TLE, the ITD is present not only between the Hilbert envelopes of the left and right ear signals, but also between the slowly varying fine structures, which may enhance the ITD representation. To test this hypothesis, a vocoder simulation experiment was conducted to see if the Bi-TLE yields better localization performance than the Bi-CIS does.

3. Methods

3.1. Subjects

Ten NH subjects (five females, ages from 20 to 25 years) participated in this study. All were students from Shenzhen University and native Mandarin speakers. Informed consent was obtained before testing in accordance with the local institution's review board. They were monetarily compensated for their participation.

3.2. Materials

The sentences from the Mandarin Hearing-In-Noise Test (MHINT) corpus [15] were used. Binaural stimuli were generated by introducing a time delay (ITD) between the two channels and then applying the carrier-synchronized vocoder simulation. Five ITDs, i.e., $-621, -337, 0, 337, \text{ and } 621 \mu\text{s}$, were derived from the head-related impulse responses from the CIPIC database [16] for horizontal azimuths of $-80^\circ, -40^\circ, 0^\circ, 40^\circ, \text{ and } 80^\circ$, respectively. Stimuli were generated in MATLAB and presented via a Roland Quad-Capture UA-55 sound card and a Sennheiser HD 650 headphone at a 16 kHz sampling rate and $\sim 65 \text{ dB A}$ in a sound-proof room.

3.3. Vocoder implementation

A 22-channel tone vocoder was used. The acoustic inputs were first divided into 22 bands by a sixth-order Butterworth band-pass filter bank, which covers the range of 80 to 7999 Hz. The corner frequencies were selected according to the Greenwood function [17], i.e., 80.0, 122.4, 172.1, 230.4, 298.7, 378.8, 472.8, 583.1, 712.3, 863.9, 1041.6, 1250.1, 1494.5, 1781.2, 2117.3, 2511.5, 2973.8, 3515.9, 4151.7, 4897.2, 5771.5, 6796.7, and 7999.0 Hz.

The envelope extraction in each channel in CIS was implemented by a half-wave rectification and an eighth-order Butterworth low-pass filter with a cutoff frequency of 250 Hz. In TLE (see Figure 4, 1) for the first three low-frequency

channels, the band-passed signals were directly used as the modulators. f_H represents the upper limit frequency of the sub-band signal and F was set to be 250 Hz. 2) for channels within the frequency range from 230.4 to 1494.5 Hz, i.e., Channel 4 to 12, the modulators were extracted by first multiplying a sinusoidal signal $\cos(2\pi f_k t)$ and then applying a low-pass filter, as shown in Figure 1(b). The value of the sinusoidal frequency f_k was set as the lower cutoff frequency of the channel minus 50 Hz (i.e., $f_k = f_{kl} - 50$). The low-pass filter was an eight-order Butterworth filter with a cutoff frequency which equals to the bandwidth of the corresponding channel plus 150 Hz. 3) for the remaining channels, i.e., Channel 13-22, the same envelope extraction process was used as in the CIS. Then, the modulator of each channel was multiplied by a sinusoidal carrier signal $\sin(2\pi f_{k0} t)$, whose frequency f_{k0} was at the center of the corresponding channel, and its initial phase was fixed at zero [12]. Finally, the modulated signals of all the channels were summed up to get a vocoded speech sound for one ear. At both ear sides, the sinusoidal carrier signals are synchronized.

3.4. Procedures

The subjects were instructed to select the perceived azimuth on a computer screen. No feedback was given. Three strategies were tested in this study, i.e., Bi-TLE, Bi-CIS and the original speech (denoted by non-vocoded; to serve as a baseline). Each strategy was tested in one block using one MHINT list (20 trials, 1 sentence/trial). All five simulated azimuths ($-80^\circ, -40^\circ, 0^\circ, 40^\circ, \text{ and } 80^\circ$, as illustrated in Figure 5) were displayed on a screen. In each trial, a sentence was presented to the listener with an ITD (corresponding to a simulated azimuth) randomly picked from the five ITDs. The subjects selected the perceived azimuth based on the azimuth sound signal heard. The order of the ITDs was randomized and counterbalanced across strategies and subjects. A training session was conducted before the formal tests to familiarize subjects with the vocoded sounds and the test procedure.

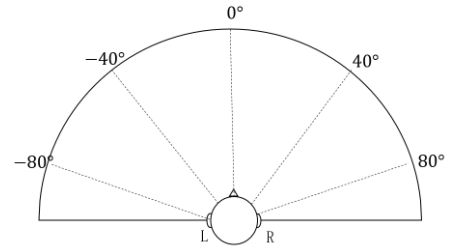


Figure 5: Five simulated azimuths using different ITDs

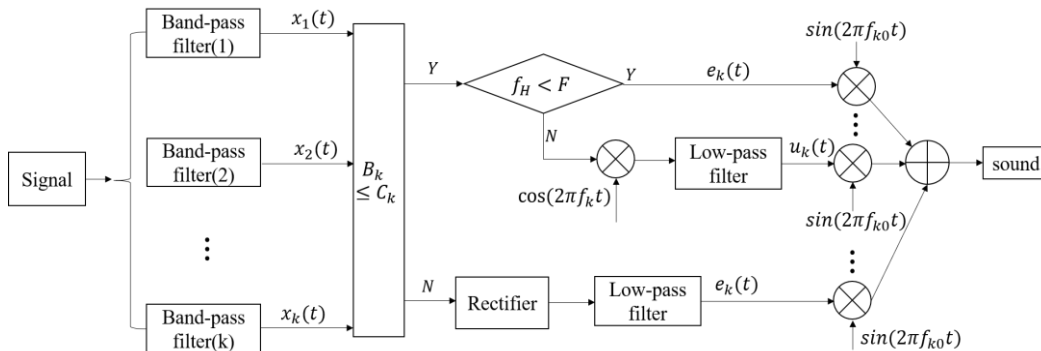


Figure 4: Functional block diagram of the vocoder simulation of the TLE strategy.

4. Results

Figure 6 shows the localization accuracy results. Subjects showed difficulty identifying sound directions using the weak envelope ITD cues presented with Bi-CIS but an overall benefit was observed. A repeated-measures analysis of variance (RM-ANOVA) revealed that there was a significant effect of test conditions. All subjects except one (Subject 5) had better performance with Bi-TLE than with Bi-CIS, although statistically the accuracy advantage of Bi-TLE over Bi-CIS was marginally significant ($p = 0.055$, pairwise comparisons with Bonferroni corrections). Both Bi-TLE ($p = 0.022$) and Bi-CIS ($p = 0.002$) vocoded conditions showed significantly poorer performance than the non-vocoded original conditions.

Detailed confusion matrices are shown in Table 1. It can be observed that for the non-vocoded stimuli, subjects had good localization performance without left-right confusion which was present in both vocoded conditions. Bi-TLE led to more accurate performance (ratios in the diagonal) compared to CIS for every azimuth. The results show that the larger the absolute azimuth, the more salient the advantage of Bi-TLE over Bi-CIS.

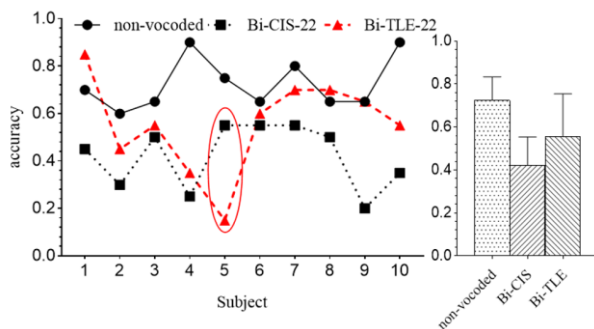


Figure 6: Sound localization accuracies: individual results (left) and group means (right; error bars: standard deviations).

Table 1: Detailed confusion matrices.

Overall ratio	Perceived azimuth (°)				
	-80	-40	0	40	80
non-vocoded simulated azimuth (°)	-80	0.575	0.425	0	0
	-40	0.325	0.65	0.025	0
	0	0.1	0.025	0.85	0.025
	40	0	0	0.15	0.625
	80	0	0	0.05	0.35
Bi-CIS simulated azimuth (°)	-80	0.275	0.575	0.125	0.025
	-40	0.05	0.4	0.5	0.05
	0	0	0.25	0.55	0.175
	40	0.05	0.025	0.3	0.475
	80	0.05	0.025	0.15	0.375
Bi-TLE simulated azimuth (°)	-80	0.525	0.325	0.1	0.05
	-40	0.1	0.5	0.225	0.125
	0	0.075	0.1	0.575	0.2
	40	0.025	0.025	0.15	0.575
	80	0.025	0	0.025	0.35

5. Discussion

This study investigates whether ITD sensitivity could be enhanced by a Bi-TLE strategy. The performance of Bi-TLE and Bi-CIS was compared in a localization test with vocoded stimuli. Results showed that listening with Bi-TLE leads to better overall performance than Bi-CIS.

This finding implies that the ITD cue with bilateral CIs may be enhanced by a Bi-TLE strategy. TLE uses a frequency transposition process to convert the fast-varying TFS into a new slowly-varying TFS, and maintained the temporal envelope. While the ITD cue can be conveyed partially between the temporal envelopes at the two ears, the TFS provides more precise ITD information. The results indicated that the CI-simulated subjects could use the more precise ITD cues provided by TLE, and thus achieved better performance.

However, the performance of TLE is much poorer than NH (the non-vocoded condition). Similar findings were reported in real Bi-CI users in the literature. This gap may be attributed to many factors such as poor spectral resolution, the difference in insertion depths of electrode arrays in the two ears, non-synchronism of the sampling clocks in the processors at the two ears, and as well as insufficient information coded in the strategy.

This study provides further data on the ITD enhancement by Bi-TLE following our previous study in [11]. In that study, Bi-TLE demonstrated a benefit in spatial release from masking than a standard envelope-based strategy. This study supplements that study in that it focuses on the ITD and evaluates the feasibility of the ITD cues. The results in this study and our previous data suggest that Bi-TLE may enhance the TFS-ITD, which at least partially contributed to the Bi-TLE benefit in spatial release from masking in [11].

There are several limitations to this study. First, there is always a gap between the vocoder simulated and real electrical hearing. The vocoder mainly simulated the coarse spectral-temporal resolution of electrical hearing, whereas other aspects were not considered (e.g., the pulsatile stimulation, the compression to fit individual dynamic range, current spread, etc.). Second, while enhanced ITD was observed in the psychophysical tasks, it is unclear whether this enhancement will provide benefits to real Bi-CI users in real-world situations where both ITDs and ILDs are present.

6. Conclusion

The Bi-TLE demonstrated overall better sound localization abilities than the Bi-CIS did in a vocoder simulation experiment. The results together with our previous experiment data in [11] and [12] indicate that the TFS-ITD cue is highly promising to be enhanced by Bi-TLE for bilateral CI users.

7. Acknowledgements

We are grateful to the listeners who participated in this experiment. This work is jointly supported by Guangdong Key Area R&D Project (No. 2018B030338001), National Natural Science Foundation of China [No. 11704129 and 61771320], and Guangdong Basic and Applied Basic Research Foundation [No.2020A1515010386]. Nengheng Zheng is the corresponding author.

8. References

- [1] R. Van Hoesel, R. Ramsden, and M. Odriscoll, "Sound-direction identification, interaural time delay discrimination, and speech intelligibility advantages in noise for a bilateral cochlear implant user," *Ear and Hearing*, vol. 23, no. 2, pp. 137-49, 2002.
- [2] R. Y. Litovsky, A. Parkinson, and J. Arcaroli, "Spatial hearing and speech intelligibility in bilateral cochlear implant users," *Ear and Hearing*, vol. 30, no. 4, pp. 419-31, 2009.
- [3] S. Kerber and B. U. Seeber, "Sound localization in noise by normal-hearing listeners and cochlear implant users," *Ear and Hearing*, vol. 33, no. 4, pp. 445-57, 2012.
- [4] R. van Hoesel, "Bilateral Cochlear Implants," *Springer handbook of auditory research*, vol. 14, no. 4, pp. 13-57, 2011.
- [5] B. Laback, K. Egger, and P. Majdak, "Perception and coding of interaural time differences with bilateral cochlear implants," *Hearing Research*, 2015, pp. 138-150.
- [6] I. Hochmair, P. Nopp, C. Jolly, M. Schmidt, H. Schöber, C. Garnham, and I. Anderson, "MED-EL cochlear implants: state of the art and a glimpse into the future," *Trends in Amplification*, vol. 10, no. 4, pp. 201-219, 2006.
- [7] 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," *Journal of The Association for Research in Otolaryngology*, vol. 19, no. 3, pp. 301-315, 2018.
- [8] 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," *Journal of the Acoustical Society of America*, vol. 143, no. 3, pp. 1428-1440, 2018.
- [9] Q. Meng, N. Zheng, and X. Li, "A temporal limits encoder for cochlear implants," *International Conference on Acoustics, Speech, and Signal Processing*, 2015, pp. 5863-5867.
- [10] 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," *Journal of the Acoustical Society of America*, vol. 139, no. 1, pp. 301-310, 2016.
- [11] A. Kan and Q. Meng, "Spatial release from masking in bilateral cochlear implant users listening to the Temporal Limits Encoder strategy," *Proceedings of the 23rd International Congress on Acoustics*, 2019.
- [12] 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," *Acoustical Science and Technology*, vol. 41, no.1, pp. 209-213, 2020.
- [13] William. A. Gardner, "Rice's Representation for Cyclostationary Processes," *IEEE Transactions on Communications*, vol. 35, no. 1, pp. 74-78, 1987.
- [14] D. Vakman, "On the analytic signal, the Teager-Kaiser energy algorithm, and other methods for defining amplitude and frequency," *IEEE Transactions on Signal Processing*, vol. 44, no. 4, pp. 791-797, 1996.
- [15] L. L. Wong, S. D. Soli, S. Liu, N. Han, and M. W. Huang, "Development of the Mandarin Hearing in Noise Test (MHINT)," *Ear and Hearing*, vol. 28, no. 2, pp. 70-74, 2007.
- [16] V. R. Algazi, R. O. Duda, and D. M. Thompson, "The CIPIC HRTF database," *Workshop on Applications of Signal Processing to Audio and Acoustics*, 2001, pp. 99-102.
- [17] D. D. Greenwood, "A cochlear frequency-position function for several species--29 years later," *Journal of the Acoustical Society of America*, vol. 87, no. 6, pp. 2592-2605, 1990.