A TEMPORAL LIMITS ENCODER FOR COCHLEAR IMPLANTS

Qinglin Meng, Nengheng Zheng, Xia Li

Shenzhen Key Laboratory of Modern Communication and Information Processing
College of Information Engineering, Shenzhen University

ABSTRACT

Cochlear implant (CI) strategies extract multi-channel temporal envelopes to stimulate an array of electrodes. However, temporal pitch perception ability is known to be limited to a low frequency range at single channels. Therefore, this study proposes a temporal limits encoder (TLE) to make full use of the temporal processing abilities at each individual channels. First the total bandwidth is allocated to uniform narrow-bands, which are then down-shifted to the CI user's low frequency range (about 50-250 Hz) of temporal pitch. After the slowly varying signals are processed by half-wave rectification, the rectified signals are compressed and used to amplitude modulate interleaved high constant rate pulse sequences. By first deriving slowly varying signals, more information or even novel information about the original signal can be preserved based on the perceptual abilities of the implantees. Preliminary psychoacoustic frequency discrimination experiments suggest that the proposed TLE strategy can offer finer frequency resolution abilities.

Index Terms— temporal cue, cochlear implants, vocoder, frequency shift, demodulation

1. INTRODUCTION

According to the NIH report, about 324,200 hearing impaired patients have been implanted with cochlear implants (CIs) by December 2012 (NIH Publication No. 11-4798). Current CI processors work as stimulating different cochlear locations with amplitude modulated pulse trains. The stimulation rate is usually set as a constant value. The temporal fluctuation in each channel is expressed by the amplitude variation of the pulses. In current clinical strategies, for example the continuous interleaved sampling (CIS), the modulation signals are the temporal envelopes derived from multi-band filters [1]. Essentially, the CIS strategy works similarly to the analyzing part of the channel vocoder [2, 3], which was widely used to reduce the transmission bandwidth of speech communication systems, e.g., from 3 kHz to 600 Hz, in around 1950s [4]. In CI products, the typical bandwidth is about 8 kHz. Although the envelope extraction processing might degrade the wideband information due to the lost of fine structures, it can be well compensated by multi-band envelopes and good speech understanding can be achieved [5]. However, CI users are still struggling with listening in the noisy conditions and perception of non-speech sounds such as music [6] and environmental sounds [7].

Several different CI strategies aiming at overcoming the abovementioned problems. Nie *et al* [8] proposed a CI strategy, namely the single sideband encoder (SSE), for better perception of music melody. In SSE, a single sideband demodulator was adopted to downward shift each sub-band signal to the base band, resulting in a slowly varying signal, termed as "coherent envelope". These envelopes were directly transmitted in analog waves or to be encoded by electric pulses at the peak time. An acoustic simulation experiment showed improved performances in melody recognition comparing with CIS.

Li *et al* [9] improved SSE in a way of "harmonic coherent demodulation" (HSSE). In SSE, the carrier frequency for the subband signal was defined as the lower cutoff frequency of the respective subband. The HSSE, however, selected an integer multiple of the fundamental frequency (*F0*), i.e., one of the harmonic frequencies, as the carrier frequency, such that the coherent envelope fluctuates synchronously with the instantaneous *F0*. Experimental studies demonstrated the improved performance of HSSE in the pitch-related perception as well as the speech perception in noisy conditions, in comparison with CIS [10, 11].

In this study, a temporal limits encoder (TLE) is proposed to make full use of the temporal processing ability at individual CI channels. Section 2 describes TLE algorithm and its advantage over CIS, SSE, and HSSE, in consideration of signal processing and psychoacoustical properties. In section 3, subjective experiments for preliminarily evaluating the frequency discrimination ability of the proposed TLE algorithm are presented, with the clinical CIS strategy as a baseline. Finally, conclusions and discussions are given in Section 4.

2. ALGORITHMS

Current CI processors stimulate different cochlear locations using interleaved pulse trains to overcome the problem of channel interference [1]. The stimulation rate is usually set as a constant value, e.g., 1000 pps. The temporal fluctuation in each channel is expressed by the amplitude variation of the pulses, with a slow variation rate (normally

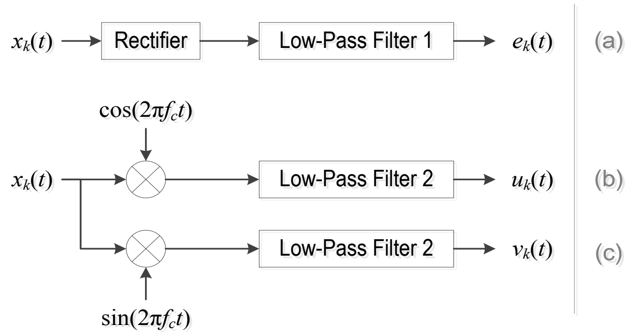

Fig. 1 Incoherent (a) and coherent (b, c) demodulations of the band-pass signal $x_k(t)$ from the k^{th} channel of a CI.

up to several hundred Hertz) [12]. The CI strategies CIS, SSE, and HSSE differ from each other in the way they extract slow varying information. This section introduces and compares each of the CI strategies as well as discusses how the proposed TLE strategy offers improvements.

2.1. Slow variation information retrieval in CIS, SSE, and HSSE

Figure 1 gives two different ways to generate a slowly varying signal from the k^{th} channel signal $x_k(t)$, namely the incoherent and coherent demodulations. The incoherent one is used in CIS, which consists of a rectifier and a low-pass filter (LPF1), as shown in Fig.1 (a). The $e_k(t)$ denotes the temporal envelope and its bandwidth is from 0 Hz to the cutoff frequency of LPF1. That is, the variation rate can be controlled by the cut-off frequency. An alternative implementation of the envelope extractor used in CI product is a Hilbert transform (HT) method [3]: the magnitude of the analytic signal corresponding to $x_k(t)$ is treated as the envelope, which can be calculated by $\sqrt{u_k^2(t)+v_k^2(t)}$ from Fig. 1 (b and c) [13].

In SSE and HSSE, a coherent demodulation is adopted to shift $x_k(t)$ to a low frequency range, as shown in Fig. 1 (b). The carrier frequency f_c is set as the lower cutoff frequency (denoted by f_l) of $x_k(t)$ in SSE, which shifts $x_k(t)$ to the base band. When the cutoff frequency of LPF2 is higher than the bandwidth of $x_k(t)$, $u_k(t)$ has the same bandwidth with $x_k(t)$. From signal processing point of view, $u_k(t)$ preserves more information comparing with $e_k(t)$ in CIS. Given the carrier, $x_k(t)$ can be recovered from $u_k(t)$, but the recovery is usually impossible from $e_k(t)$. What's more, SSE usually gives better frequency discrimination than CIS. For example, two sinusoidal signals with different frequencies in $x_k(t)$ still correspond to two different sinusoidal signals in $u_k(t)$, while in CIS they usually produce similar $e_k(t)$, especially when $x_k(t)$ is in the high frequency range.

However, the SSE shifts frequency components near f_l to very low frequency, e.g., 0-50 Hz, which is rarely perceived as pitch variations by CI users [12, 14]. Furthermore, the bandwidths in high frequency channels of SSE are much higher than 300 Hz, which was generally suggested as the upper limit for the temporal pitch at single site [15]. As a result, the benefits of frequency down-shifting in SSE cannot be fully obtained.

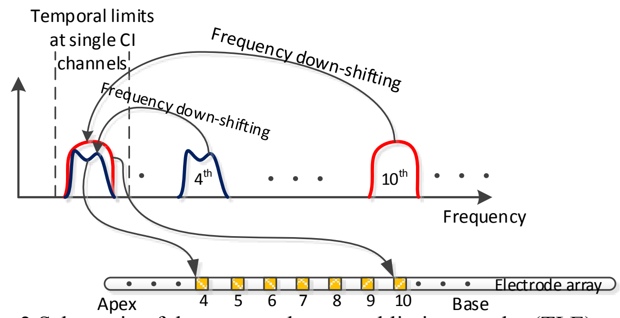

Fig. 2 Schematic of the proposed temporal limits encoder (TLE)

In HSSE, the frequency of the strongest harmonic component in $x_k(t)$, denoted as $mF0$, is first detected. Then, f_c is set as $(m\text{-}1)F0$, which ensures that $u_k(t)$ from multi channels fluctuate synchronously in a temp of $F0$ or low integer multiples of $F0$ [9]. Pitch perception was found improved by HSSE, comparing to CIS [11]. However, no benefit on non-harmonic sound perception can be achieved from HSSE, and harmonic component detection usually fails in noisy environments. On the other hand, the same frequency component from two different sounds may be transmitted by HSSE to different channels due to concurrent other frequency components. This limits the application of HSSE for current CI processors.

2.2 Slow variation information retrieval in TLE

A schematic of the proposed TLE strategy is given in Fig. 2, we adopt the same processing as in SSE for frequency down-shifting, i.e., Fig. 1 (b), except that the carrier frequency is suggested to be slightly lower than f_l, e.g., f_l–50 Hz. In this way, the frequency components above f_l are shifted to a frequency range higher than 50 Hz and the temporal pitch perception can be improved [14].

The cochlea works as a Fourier analyzer to analyze the incoming sound to different frequencies tonotopically, with a frequency resolution monotonically increasing from base to apex. This pattern was mathematically modeled (Greenwood function [16]) based on extensive data from physiological measurements.

In CIS, SSE, and HSSE, the CIs were expected to give a non-uniform frequency resolution similar to human cochlea. However, studies found that temporal electric pitch perception at single electrode site was uniformly limited to a few hundred Hertz [15, 17], which was not found to be wider at more-basal electrodes. This fact limits the total bandwidth utilization of current CIs, which have only 12-24 channels and usually cannot reach the most apical turn of the cochlea.

In TLE, the frequency range allocated to each channel should be narrower than the temporal limits frequency range (about 50-300Hz) reported at the neural interfaces [15]. This study selects a bandwidth of 200Hz, so for example the total bandwidth of a 16-channel CI device would be 3.2kHz,

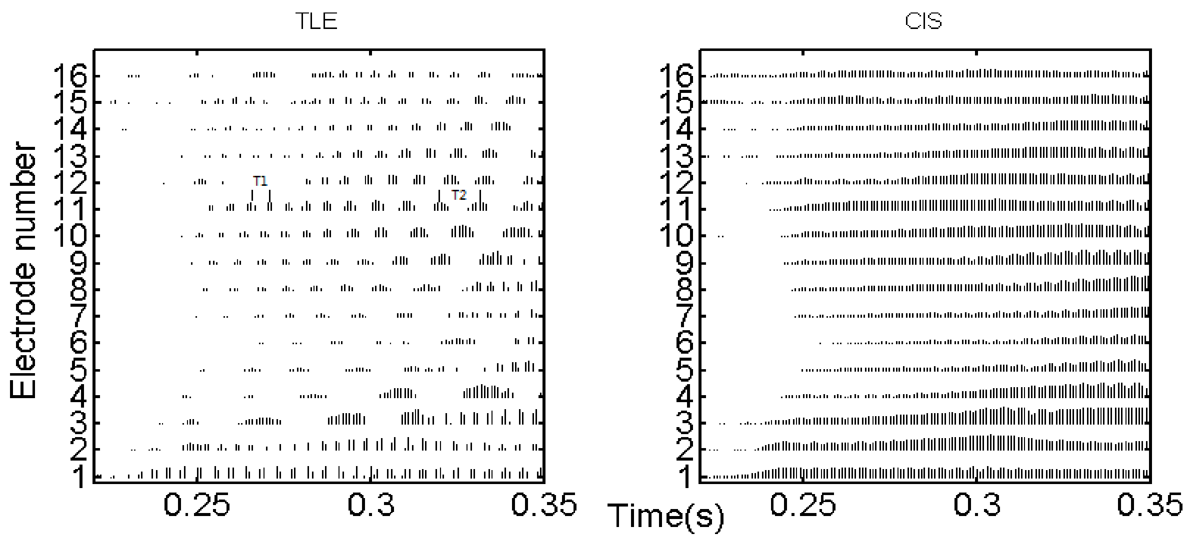

Fig.3 Electrodogram examples from temporal limits encoder (TLE) and continuous interleaved sampling (CIS) strategies.

rather than the conventional 8 kHz. According to the down-shifting method of TLE, each bandpass signal is shifted to 50-250 Hz, which is the range of temporal pitch ranking [12, 14]. In fact, telephony sounds cover only from approximately 300 to 3,400 Hz [18], which is known to be sufficient for speech communication. These limits were imposed by the low channel capacity of old communication systems. Likewise, the principles of the TLE settings are based on presenting narrowband sounds with high fidelity rather than wideband sounds with low quality.

Moreover, an alternative configuration is also recommended in TLE. Firstly, the frequency allocation still acts in accordance with the non-uniform Greenwood function mode. Then, the output from each band-pass filter is down-shifted with f_l–50. Then, for the channels whose bandwidth is narrower than 200Hz, the down-shifted signals are defined as the slow variation information, and for the other channels, incoherent envelopes of the down-shifted signals are extracted as Fig.1 (a) in CIS.

2.3. Acoustic-to-electric mapping

Acoustic-to-electric mapping refers to the process that converts the slowly varying signals from each acoustic channel, i.e., $e_k(t)$ in CIS or $u_k(t)$ in SSE, HSSE, and TLE, to the electric pulse trains at the individual electrodes. The mapping mechanism in CIS uses pulse carriers with constant rate to sample the $e_k(t)$. While SSE proposed to first detect the positive peaks of $u_k(t)$ and pulses are generated in synchrony with the peaks. And in HSSE, the $u_k(t)$ is first band-pass filtered (with bandwidth of 50 - 300 Hz) and then half-wave rectified, and at last the output is sampled with a constant rate pulse carrier. In the proposed TLE, the $u_k(t)$ is first half-wave rectified and then sampled with a constant rate pulse carrier. The half-wave rectification is similar to the sound processing mechanism of the basilar membrane and the upper and lower envelops are symmetrical for the narrow band signals [19]. For all strategies, interleaved sampling is adopted among channels. The compression function can be logarithmic or power-law [20].

2.4 Brief summary

From signal processing point of view, the TLE suggests a lower carrier frequency (f_l–50 Hz) and a uniform bandwidth allocation with a narrow bandwidth for each channel. These configurations ensure that different frequencies, in a continuous speech frequency range (e.g., about 50-3250 Hz for 16-channel condition), are expressed as different amplitude periodicities at certain channels, which, according to the psychophysical knowledge, can be ranked in a pitch scale. This property is not available in CIS, SSE, and HSSE as discussed in Section 2.1. The proposed TLE, which is an extension of SSE strategy, is proposed to improve the performance of CI. In this preliminary study of TLE, the basic temporal and spectral perceptions are investigated in Section 3, comparing with the CIS strategy.

3 PRELIMINARY EVALUATIONS

3.1 Temporal representation

The temporal envelope of $u_k(t)$ is almost identical with that of $x_k(t)$ [13, 21], which means that the envelope in TLE preserves the general trend of temporal envelope in CIS. Moreover, $u_k(t)$ preserves more fine structure than CIS envelops. For example, Fig. 3 illustrates the electrodograms from TLE and CIS for a middle section of a Chinese word "下" (/xià/) pronounced by a male. Here, in order to directly compare $u_k(t)$ and $e_k(t)$, both TLE and CIS used uniform bandwidth allocation. The frequency range of the k^{th} channel is from $200k$ – 150 Hz to $200k$ + 50 Hz. The lexical tone (Tone 4, i.e., the falling one) can be observed from the TLE electrodogram through the envelope periodicity in multi channels (e.g., T2>T1 in Fig.3), but not from the CIS electrodogram. More information about acoustic correlates

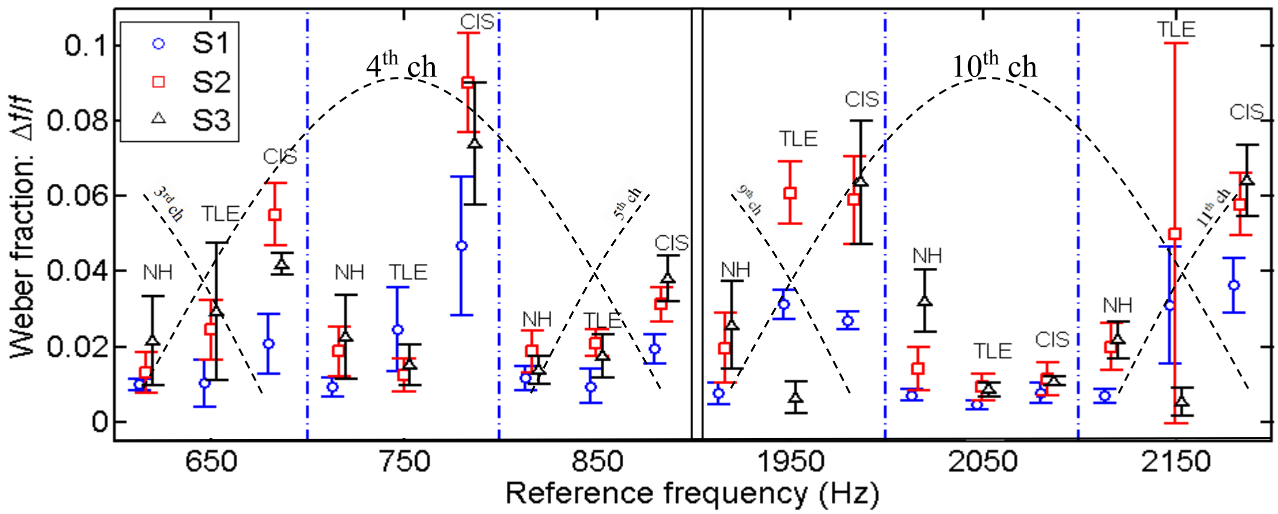

Fig.4 Frequency difference limens for normal hearing (NH), temporal limits encoder (TLE) simulation, and continuous interleaved sampling (CIS) simulation. The dashed curves here only represent the relations between different reference frequencies in each channel.

of Chinese tone perception with CIs can be referred to some acoustic papers, e.g., [22] and [23].

3.2 Frequency discrimination

To evaluate the frequency discriminating abilities of proposed TLE strategy, pure tone frequency detection difference limen was measured on normal hearing, TLE, and CIS. A noise carrier acoustic simulation model [5] was adopted. Three young normal hearing (NH) subjects participated (S1, S2, and S3). The stimuli were presented through a DUO-CAPTURE USB audio interface and a Sennheiser HD650 headphone at a comfortable level in an anechoic chamber. The stimuli were pure tones around 6 different reference frequencies (650, 750, 850, 1950, 2050, and 2150 Hz), where 750 and 2050 Hz were the center frequencies of the 4th and 10th channel, and the other 4 frequencies were the cutoff frequencies.

A standard psychoacoustical procedure, i.e., a three-down, one-up, three-alternative, forced-choice, adaptive procedure, was adopted for the evaluation. Three listening conditions were evaluated, i.e., listening to the unprocessed pure tones (NH), the TLE vocoder processed pure tones, and the CIS vocoder processed pure tones. In the experiments, the order of reference frequencies and listening conditions were randomized and the results averaged the best 4 results from 6 estimates.

The frequency discrimination results are illustrated as Weber fractions, i.e., $\Delta f/f$, where f is the reference frequency and Δf is the difference limen, in Fig. 4. The error bars indicate the 95% confidence intervals. As can be seen, when the reference frequencies are 650, 750, 850, and 2050 Hz, all subjects showed significantly lower limens with TLE than with CIS (t-test: $p < 0.05$), which tells the better frequency discriminating power of TLE. When the reference frequencies are 1950 and 2150 Hz, TLE showed significant advantage over CIS for S3 ($p < 0.00005$), while for S1 and S2 TLE and CIS are comparable ($p > 0.38$) except for S1 with 1950 Hz ($p = 0.024$) showing a marginally negative evidence. Under some conditions, the means of TLE are comparable to or even lower than NH, while CIS generally showed higher limens than NH. These results suggest that the TLE preserves better the frequency information both at the center and the corners of certain channels of a cochlear implant than the CIS.

4. CONCLUSIONS AND DISCUSSIONS

Inspired from the temporal limits at the electrode-to-nerve interface, the TLE was proposed to make full use of the temporal processing ability at each channel. Unlike the envelope extraction operation in CIS, TLE downshift the uniformly allocated, fast-varying bandpass signal to the temporal limits range (about 50-250 Hz), such that more pitch related temporal information can be preserved. Preliminary experiments of vocoder simulation on normal hearing subjects were carried out to compare the frequency discrimination ability between TLE and CIS. Results suggested that TLE might provide finer frequency discrimination ability to implantees than CIS. Experiments to explore the CI users' adaptation to the novel temporal information will be conducted in the next stage.

5. ACKNOWLEDGEMENTS

Thank Xiaowei Zhang and Dujuan Liu for their participation in the psychoacoustic experiment. Thank the anonymous referees who provided valuable suggestions leading to significant improvements on the manuscript writing. This study was jointly supported by Shenzhen Scientific Research & Development Fund (JCYJ20120614085245889), a Shenzhen key laboratory project (CXB201105060068A), and a fund awarded by China Scholarship Council (201308440223).

6. REFERENCES

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

[2] H. Dudley, "Remaking Speech," *J Acoust Soc Am,* vol. 11, pp. 169-177, 1939.

[3] P. C. Loizou, "Speech processing in vocoder-centric cochlear implants," *Adv Otorhinolaryngol,* vol. 64, pp. 109-143, 2006.

[4] B. Gold and C. M. Rader, "The channel vocoder," *IEEE Transactions on Audio and Electroacoustics,* vol. 15, pp. 148-161, 1967.

[5] R. V. Shannon, F. G. Zeng, V. Kamath, J. Wygonski, and M. Ekelid, "Speech recognition with primarily temporal cues," *Science,* vol. 270, pp. 303-303, 1995.

[6] C. J. Limb and J. T. Rubinstein, "Current research on music perception in cochlear implant users," *Otolaryngol Clin North Am,* vol. 45, pp. 129-40, 2012.

[7] V. Shafiro, B. Gygi, M. Y. Cheng, J. Vachhani, and M. Mulvey, "Perception of environmental sounds by experienced cochlear implant patients," *Ear Hear,* vol. 32, pp. 511-23, 2011.

[8] K. Nie, L. Atlas, and J. Rubinstein, "Single sideband encoder for music coding in cochlear implants," *in Proc. ICASSP,* Las Vegas, USA, 2008, pp. 4209-4212.

[9] X. Li, K. Nie, L. Atlas, and J. Rubinstein, "Harmonic coherent demodulation for improving sound coding in cochlear implants," in *Proc. ICASSP,* Dallas, USA, 2010, pp. 5462-5465.

[10] X. Li, K. Nie, N. S. Imennov, J. H. Won, W. R. Drennan, J. T. Rubinstein*, et al.*, "Improved perception of speech in noise and Mandarin tones with acoustic simulations of harmonic coding for cochlear implants," *J Acoust Soc Am,* vol. 132, pp. 3387-3398, 2012.

[11] X. Li, K. Nie, N. Imennov, J. Rubinstein, and L. Atlas, "Improved Perception of Music with a Harmonic Based Algorithm for Cochlear Implants," *IEEE Trans Neural Syst Rehabil Eng,*vol. 21, pp. 684-494, 2013.

[12] C. M. McKay, H. J. McDermott, and G. M. Clark, "Pitch percepts associated with amplitude-modulated current pulse trains in cochlear implantees," *J Acoust Soc Am,* vol. 96, pp. 2664-73, Nov 1994.

[13] D. Vakman, "On the analytic signal, the Teager-Kaiser energy algorithm, and other methods for defining amplitude and frequency," *IEEE Trans Sig Proc,* vol. 44, pp. 791-797, Apr 1996.

[14] Q. Meng, M. Yuan, H. Mou, and H. Feng, "Envelope pitch at different stimulation sites of cochlear implant," *J Acoust Soc Am,* vol. 131, pp. 3517-3517, 2012.

[15] F. G. Zeng, "Temporal pitch in electric hearing," *Hear Res,* vol. 174, pp. 101-6, Dec 2002.

[16] D. D. Greenwood, "A cochlear frequency - position function for several species - 29 years later," *J Acoust Soc Am,* vol. 87, pp. 2592-2605, 1990.

[17] G. M. Clark, "The multi-channel cochlear implant: Multi-disciplinary development of electrical stimulation of the cochlea and the resulting clinical benefit," *Hear Res,* Available online, in-press, Aug 24 2014.

[18] H. Pulakka and P. Alku, "Bandwidth Extension of Telephone Speech Using a Neural Network and a Filter Bank Implementation for Highband Mel Spectrum," *IEEE Trans Audio Speech and Language Processing,* vol. 19, pp. 2170-2183, Sep 2011.

[19] Q. Meng, M. Yuan, J. Zhao, and H. Feng, "Experimental study on rationality of 'Hilbert envelope'based on Empirical Mode Decomposition," in *Proc. Int. Conf. Audio, Language and Image Processing (ICALIP),* Shanghai, China, 2012, pp. 616-620.

[20] P. C. Loizou, "Mimicking the human ear," *IEEE Signal Processing Magazine,* vol. 15, pp. 101-130, 1998.

[21] W. Gardner, "Rice's representation for cyclostationary processes," *IEEE Trans Commun,* vol. 35, pp. 74-78, 1987.

[22] Q. J. Fu, F. G. Zeng, R. V. Shannon, and S. D. Soli, "Importance of tonal envelope cues in Chinese speech recognition," *J Acoust Soc Am,* vol. 104, pp. 505-10, Jul 1998.

[23] Y. Y. Kong and F. G. Zeng, "Temporal and spectral cues in Mandarin tone recognition," *J Acoust Soc Am,* vol. 120, pp. 2830-40, 2006.