J. Biomed. Eng. Res : 417-426, 2010

Improved Melody Recognition Performance of a Cochlear Implant Speech Processing Strategy Using Instantaneous Frequency Encoding Based on Teager Energy Operator

Sung Jin Choi, Sang Baek Ryu, Kyung Hwan Kim

Department of Biomedical Engineering, College of Health Science, Yonsei University, 234 Maeji-ri, Heungup-myun, Wonju, Kangwon-do, 220-710, South Korea

(Received September 1, 2010. Accepted November 21, 2010)

Abstract

We present a speech processing strategy incorporating instantaneous frequency (IF) encoding for the enhancement of melody recognition performance of cochlear implants. For the IF extraction from incoming sound, we propose the use of a Teager energy operator (TEO), which is advantageous for its lower computational load. From time-frequency analysis, we verified that the TEO-based method provides proper IF encoding of input sound, which is crucial for melody recognition. Similar benefit could be obtained also from the use of a Hilbert transform (HT), but much higher computational cost was required. The melody recognition performance of the proposed speech processing strategy was compared with those of a conventional strategy using envelope extraction, and the HT-based IF encoding. Hearing tests on normal subjects were performed using acoustic simulation and a musical contour identification task. Insignificant difference in melody recognition performance was observed between the TEO-based and HT-based IF encodings, and both were superior to the conventional strategy. However, the TEO-based strategy was advantageous considering that it was approximately 35% faster than the HT-based strategy.

Key words : Cochlear implant, speech processing, strategy melody recognition, instantaneous frequency, Teager energy operator

Ⅰ. INTRODUCTION

Cochlear implant (CI) has been extensively used to for the restoration of hearing function of totally deaf patients due to sensorineural hearing loss. The hearing function is restored by direct stimulation of auditory nerves by electrical pulses, which are generated by speech processing strategy from incoming sound. Speech processing strategy, which is also called coding strategy, means the method for modulating stimulus pulses based on incoming sound. Basically it is rooted from the function of the peripheral auditory system. Most speech processors of current CI devices utilize bandpass filter arrays for the frequency decomposition of incoming sound, which mimics the frequency decomposition function of basilar membrane in biological cochlea [1]-[3]. While the CIs can provide successful hearing function in quiet environments, their performance rapidly deteriorates in the presence of background noise. Besides, considering poor performance in music perception, speaker identification, and sound localization, there remains much to be done to enhance the performance of CI speech processing strategy.

Corresponding Author : Kyung Hwan Kim
Department of Biomedical Engineering, College of Health Science, Yonsei University
234 Maeji-ri, Heungup-myun, Wonju, Kangwon-do, 220-710, South Korea
Tel : +82-33-760-2364 / Fax : +82-33-763-1953
E-mail : khkim0604@yonsei.ac.kr

The envelope of speech waveform is known to be the most important perceptual cue, and thus it is primary information utilized in most current CI devices. However, for more complete understanding of incoming sound, it is recognized that other information should also be incorporated in the speech processing for CI. In general, signals can be separated into slowly-varying envelope and fast-varying detail, which is called fine structure [4]. It is accepted that the fine structure plays very important roles in several applications such as speech perception in noise, pitch perception, and sound localization [4]-[5]. Thus, several approaches have been suggested for the CI speech processor that exploits the fine structure information in addition to the envelope [6]-[10]. Many of them focused on the use of instantaneous phase as

information on fine structure [6]-[8], and adopted Hilbert transform (HT). Zeng et al. [9]-[10] tried the encoding of instantaneous frequency (IF) for the incorporation of fine structure in CI speech processor, and showed considerable improvement of speech perception performance under competing voice. They also utilized the HT for the IF encoding, although they also suggested a novel signal processing scheme for the extraction of IF [9]-[10].

Although the effectiveness of the use of fine structure encoding was recognized and also experimentally demonstrated, some issues should be resolved before more widespread application of the fine structure encoding to CI speech processors. For example, it is rather uncertain how to modulate the stimulus pulses so that the fine structure cues are properly encoded in neural activities of auditory nerves. Although the usefulness of fine structure encoding using the HT was experimentally verified, additional computational load resulting from the use of HT may not be appropriate for the CI speech processor considering the requirement of real-time processing.

A Teager energy operator (TEO) is an alternative algorithm that can be used to extract the fine structure of a signal [12]. The name has originated from its capability of tracking the energy of a linear oscillator, and its output corresponds to the square of the product of signal's instantaneous amplitude (i.e. envelope) and frequency (i.e. fine structure). Maragos et al. proposed several TEO-based methods for the extraction of instantaneous amplitude and frequency [13]. They compared the TEO-based methods with conventional ones using the HT, and showed that the TEO can provide similar performance of the extraction of instantaneous amplitude and frequency [14].

Based on this, we propose to adopt the TEO-based IF extraction for the speech processor of CI. We expected that we can achieve the benefit of fine structure encoding at much lower computational cost, so that it is better suited for the real-time signal processing, and low power consumption. We focused on finding out whether the TEO-based method yields similar or superior performance for melody perception, compared to the HT-based method. The melody perception is known to be benefited greatly from the use of the fine structure. We performed hearing tests on normal-hearing subjects using acoustic simulation employing a musical contour identification task [15] in order to show the benefit of using the IF extracted by the TEO-based method.

II. METHODS

A. Proposed Speech Processing Strategy

Figure 1 (a) shows the general structure of speech processors in a CI device. In actual application, the auditory nerves are stimulated by electrical pulses generated from the speech processor as denoted by the solid lines in Figure 1 (a). The dashed lines in Figure 1 (a) indicate the procedure of generating output waveforms for acoustic simulation of the speech processor.

Figure 1 (b) illustrates the details of conventional speech processing strategies [1]-[3]. Incoming speech is decomposed into multiple subbands by an array of bandpass filters. Output signals from each subband are given to half-wave rectifiers and low pass filters for the envelope extraction. In actual application, the amplitudes of stimulation pulses at each channel are modulated by the extracted envelopes, and then these pulse trains are delivered to the electrode array for the stimulation of the auditory nerve.

The proposed speech processing strategy shown in Figure 1 (c) includes the bandpass filter array just as the conventional strategy, however, it also exploits the fine structure of incoming sound in addition to the envelope [9]-[10]. From the subband signals, instantaneous frequencies are extracted as well as the envelopes, i.e., the instantaneous amplitudes. The method for instantaneous amplitude and frequency extraction using TEO is described in detail below in section 2.2. Here the conventional envelope extractor that consists of a rectifier and a lowpass filter is eliminated since by the TEO-based method can provide identical results for the envelope extraction.

B. Instantaneous Amplitude and Frequency Extraction Based on TEO

A narrowband signal $x(t)$ can be expressed in terms of the instantaneous amplitude $A(t)$ and the IF $f_i(t)$ as follows:

$$x(t) = A(t)\cos\left[2\pi f_c t + 2\pi\int_0^t f_i(\tau)d\tau + \theta_c\right], \qquad (1)$$

where f_c and θ_c are center frequency and initial phase respectively. When the HT is used for the extraction of $A(t)$ and $f_i(t)$ out of $x(t)$, an analytical signal for $x(t)$, $z(t)$, is derived as follows:

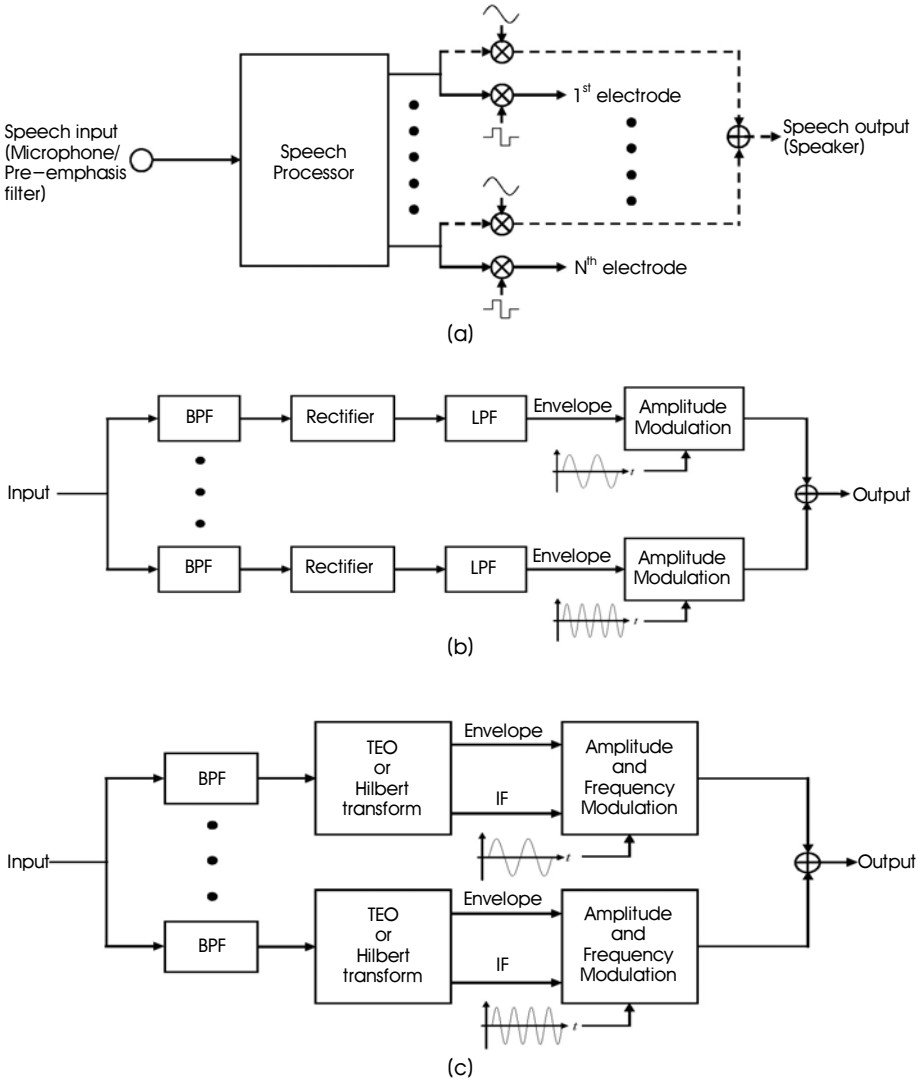

Fig. 1. (a) General structure of speech processors in a CI device. (b) Detailed block diagram of a conventional speech processing strategy. (c) Detailed block diagram of the proposed speech processing strategy employing additional IF encoding by the TEO.

$$z(t) = x(t) + j\hat{x}(t) = A(t)e^{j\theta(t)} \tag{2}$$

Here $\hat{x}(t)$ denotes the HT of $x(t)$ which is obtained from the convolution of $x(t)$ and $1/\pi t$ [11]. Obviously $A(t)$ is readily derived from $z(t)$, and $f_i(t)$ can also be obtained from the time derivative of $\theta(t)$ as shown follows:

$$r(t) = \sqrt{x^2(t) + \hat{x}^2(t)} \approx |A(t)|, \tag{3}$$

$$f_i(t) = \frac{1}{2\pi}\dot{\theta}(t) = \frac{1}{2\pi}\frac{d}{dt}\left[\tan^{-1}\left(\frac{\hat{x}(t)}{x(t)}\right)\right], \tag{4}$$

For the computation of the HT in Equation 2, the convolution of $x(t)$ and $1/\pi t$ should be obtained for all time samples, and it may cause heavy computational load, even if it is usually implemented using fast convolution, based on frequency-domain calculation by fast Fourier transform (FFT) [14].

The TEO-based extraction of instantaneous amplitude and frequency is based on the fact that the TEO can estimate the squared product of the amplitude and frequency of a narrowband signal as follows [12], [13]:

$$\Psi_c[x(t)] = \dot{x}^2(t) - x(t)\ddot{x}(t) \approx \left[A(t)f_i(t)\right]^2, \tag{5}$$

It can be shown that the instantaneous amplitude and frequency are calculated from the outputs of TEO applied to the signal $x(t)$ and its time derivative $\dot{x}(t)$ as follows [13]:

$$\frac{\Psi_c[x(t)]}{\sqrt{\Psi_c[\dot{x}(t)]}} \approx |A(t)|, \tag{6}$$

$$\sqrt{\frac{\Psi_c[\dot{x}(t)]}{\Psi_c[x(t)]}} \approx f_i(t), \tag{7}$$

When it is implemented using sampled discrete-time signal, the discrete-time counterpart of the TEO is used instead, which is expressed as

$$\Psi_d[x(n)] = x^2(n) - x(n-1)x(n+1) \approx [A(n)f_i(n)]^2, \tag{8}$$

Since only elementary numerical operation such as squaring, multiplication, and addition of a few time-samples are involved in the calculation, the TEO-based method is much more efficient in terms of computational load than the HT-based method. Thus it is better suited for the real-time processing of CI speech processor.

C. Test Materials and Acoustic Simulation

We adopted a melody contour identification (MCI) task [15] for the evaluation of the melody recognition performance of speech processors using acoustic simulation. The MCI task was originally proposed to quantify the melody recognition capability of the CI recipients [15]. Figure 2 (a) illustrates all six types of melodic contours employed in this study. Each melodic contour consists of 5 notes of equal durations. The frequencies of the 5 notes were changed to construct melodic contours of the shapes illustrated in Figure 2 (a). First, the frequency of the lowest frequency note in the contour, f_{ref}, which we call root note, was selected. And then, all the other notes in the contours were generated so that their frequency f satisfied $f = 2^{\frac{x}{12}} f_{ref}$, where x is the number of semitones, relative to the root note. As was performed by Galvin et al. [15], three different frequencies of the root notes (220 Hz, 440 Hz and 880 Hz) were employed for the performance evaluation at different frequency ranges. The number of semitones between two successive notes was changed to control the degree of difficulty for the identification: one semitone (difficult), three semitones (intermediate) and five semitones (easy).

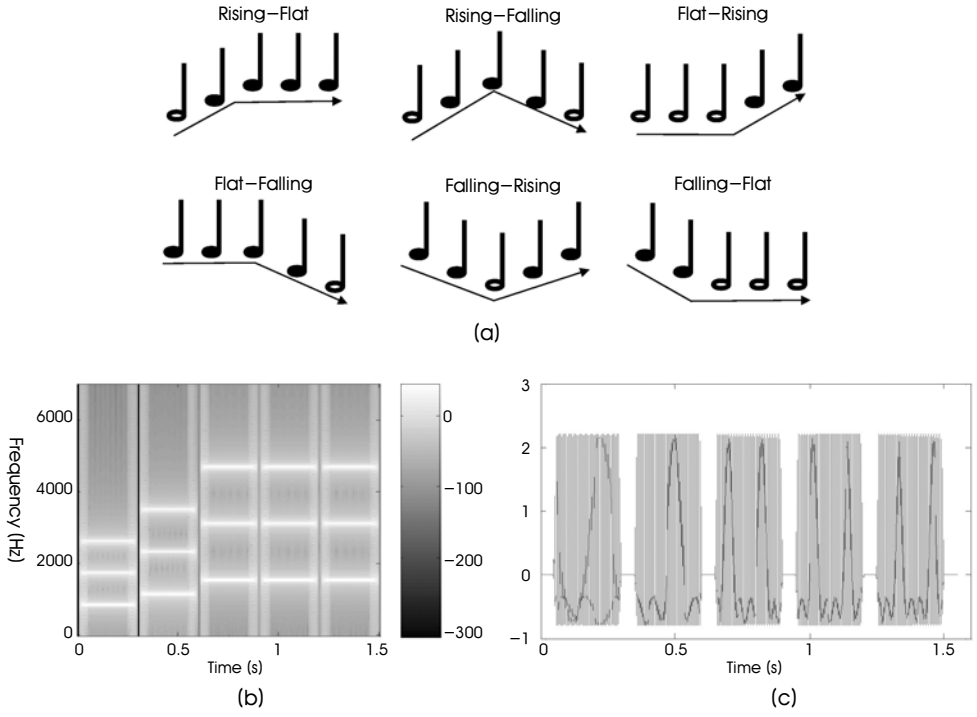

Fig. 2. (a) Six types of melodic contours employed for the hearing test. Each of the melodic contours consists of 5 notes of equal durations. (b) The spectrogram of an example of the, shown as the first example in Figure 2 (a). (c) The temporal waveform of the "Rising-Flat" melodic contour.

Figure 2 (b) and (c) show the spectrogram and waveform of an example of the "Rising-Flat" contour, shown as the first example in Figure 2 (a). The root note of this example contour in Figure 2 (b) and (c) was 880 Hz and the number of semitones between successive notes was five. Every note was 250 ms long in duration, including 10 ms onset and offset periods with linearly increasing and decreasing frequencies at the beginning and end. The interval between two successive notes was set to 50 ms.

For the evaluation of speech processing strategies for melody recognition, hearing experiments on normal subjects were performed using acoustic simulation of the proposed and conventional speech processing strategies. We adopted the synthesis of multiple sinusoids for the acoustic simulation, similarly to Zeng et al. [9]. Acoustic waveforms were generated from the waveforms of melodic contours so that they corresponded to the outputs of conventional and the proposed speech processors. For the proposed strategy, the instantaneous amplitude and frequency of each sinusoid were modulated according to the instantaneous amplitude and frequency of the waveforms at each channel of the bandpass filter arrays in Figure 1 (c). The extraction of the IF was performed by the TEO for the proposed strategy. We also implemented the IF extraction using the HT for comparison. As shown in Table 1, the center frequencies of each bandpass filter are distributed according to logarithmic scale, which was motivated by tonotopical organization of human basilar membrane [3]. The number of channels was varied to 4, 8, and 12. For the synthesis of acoustic waveforms for the conventional strategy, only the amplitudes of sinusoids were modulated by instantaneous amplitudes (i.e. envelopes). The cutoff frequency of the lowpass filter for the envelope extraction in the conventional speech processor (Figure 1 (b)) was set to 500 Hz [3].

Twelve subjects with normal hearing capability were participated in the study. All of them were paid for the participation in the experiment. 162 sound tokens were presented to each patient (6 melodic contour types×3 root notes×3 difficulty levels×3 strategies). The stored sounds were played by clicking icons in a graphic user interface, and presented binaurally using a headphone (Sennheiser HD25SP1) and a 16-bit sound card (SoundMAX^{TM} integrated digital audio soundcard). The order of presenting each sound token was randomized. The sound level was controlled to be comfortable for each subject (range: 70-80 dB, approximately). A 5-min training session was given before the main experiment. After hearing each sound, the subjects were requested to choose the correct contour corresponding to the presented one, among six types of the melodic contours of Figure 1 (a). The percentage of correct answers was scored.

Ⅲ. RESULTS

A. The Time-frequency Analysis

Figure 3 shows an example of applying the proposed TEO-based method for the extraction of envelope and IF. A

Table 1. The Center Frequencies and the Bandwidths of Each Channel of the Bandpass Filter Array.

(a) 4 Channel Implementation

Channel	1	2	3	4
Center Frequency	460	953	1971	4078
Bandwidth	321	664	1373	2842

(b) 8 channel Implementation

Channel	1	2	3	4	5	6	7	8
Center Frequency	394	692	1064	1528	2109	2834	3740	4871
Bandwidth	265	331	431	516	645	805	1006	1257

(c) 12 channel Implementation

Channel	1	2	3	4	5	6	7	8	9	10	11	12
Center Frequency	274	453	662	905	1190	1521	1908	2359	2885	3499	4215	5050
Bandwidth	165	193	225	262	306	357	416	486	567	661	771	900

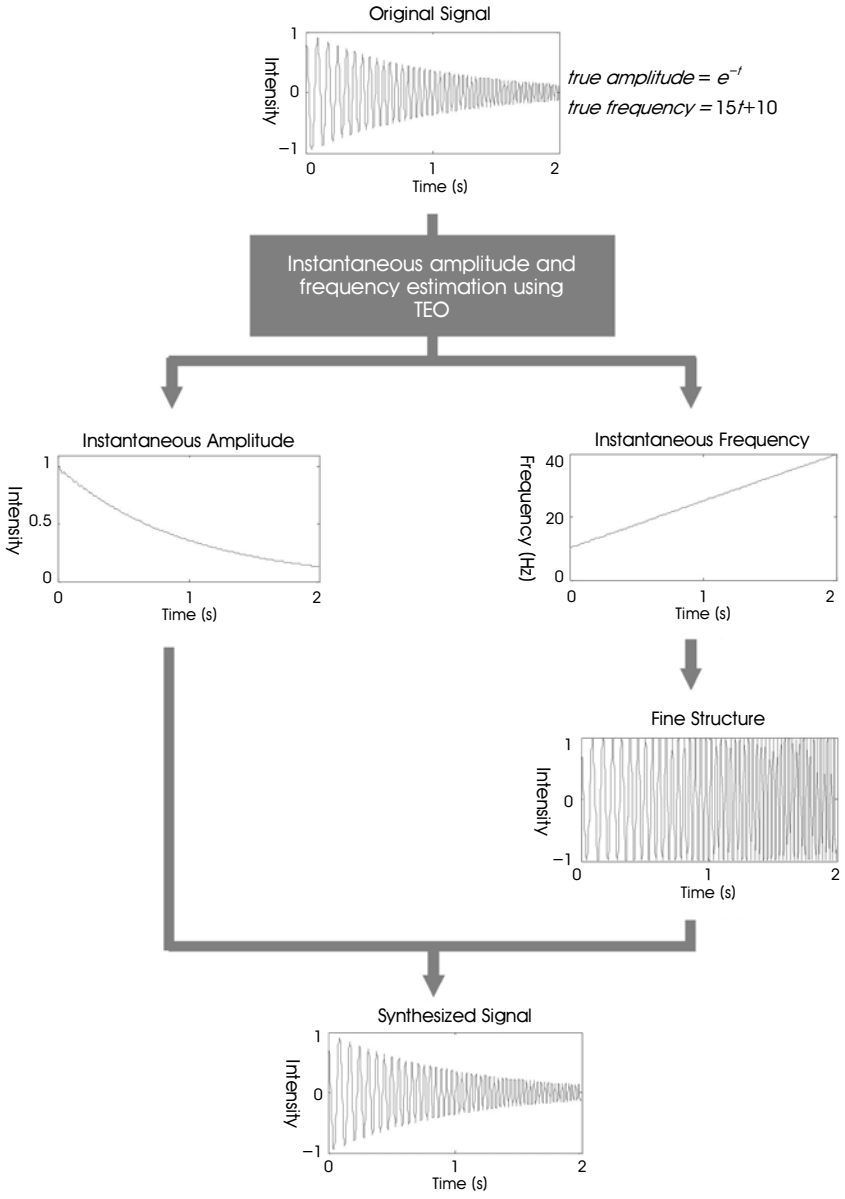

Fig. 3. An example of applying the TEO-based envelope and IF extraction to a chirp signal.

chirp signal was generated so that instantaneous frequency was linearly increased from 15 Hz to 40 Hz during 2 second period, and the envelope was exponentially decreased. The instantaneous amplitude and frequency of original input signal could be extracted faithfully as shown in Figure 3.

The representation of frequency contents at the outputs of the CI speech processors were compared by the inspection of the time-frequency activation patterns. The input sound was a rising-falling melodic contour (the 2nd example in Figure 2 (a)) with three semitones between two successive notes, and its spectrogram is shown in Figure 4 (a). Figure 4 (b) and (c) show the spectrograms of the output from the conventional and the proposed speech processing strategies, respectively (8 channels). Figure 4 (d) is the spectrogram of the output from the proposed structure, but here the envelope and IF extraction was performed by the HT instead. From Figure 4 (b), when the conventional strategy was used, it was evidently not possible to properly represent the temporal variation of instantaneous frequency, which is crucial for the melody recognition. Comparing the two strategies employing the IF encoding, regardless of using the TEO or HT, both yielded satisfactory representation of frequency contents of the input sound, as

both spectrograms in Figure 4 (c) and (d) were very similar to the spectrogram of original input shown in Figure 4 (a).

B. MCI Task

Figure 5 shows the melody recognition performances of the conventional strategy and the two strategies employing the IF encoding. The results shown in Figure 5 were obtained from acoustic simulation and hearing experiment using the MCI task. For the 4 channel implementation shown in Figure 5 (a), the two strategies using IF encoding yielded better melody recognition performance than the conventional strategy, regardless of task difficulty. Their superiority was also statistically significant (one-way ANOVA, $p<0.01$). Post-hoc pairwise t-test showed that there was no substantial difference in melody recognition performance between the two speech processors employing either TEO or HT for the IF encoding (t-test, difficult: $p=0.071$, moderate: $p=0.999$, easy: $p=0.166$). For the 8 channel and 12 channel speech processors (Figure 5 (b) and (c), respectively), statistically significant enhancement in melody recognition by additional IF information could be obtained in the difficult task (t-test, $p<0.01$).

C. Computational Load

For the comparison of computational loads of the two processing strategies using the IF encoding, we measured computation time for the generation of acoustic waveforms. The input was a sinusoidal signal with 10 second duration. Figure 6 shows the comparison of the computational times of the HT-based, and the TEO-based strategies, for the IF encoding. For 16 channel implementation, the average computation time obtained from 40 repetitions was 9.86 s for the TEO-based speech processor, whereas it was 15.57 s for the Hilbert transform-based strategy, and thus, the former was faster than the latter by 36.65% in terms of computation time (Matlab implementation in a PC with Intel Core 2 Duo Processor with 1.86 GHz clock, and 2 GB RAM). The

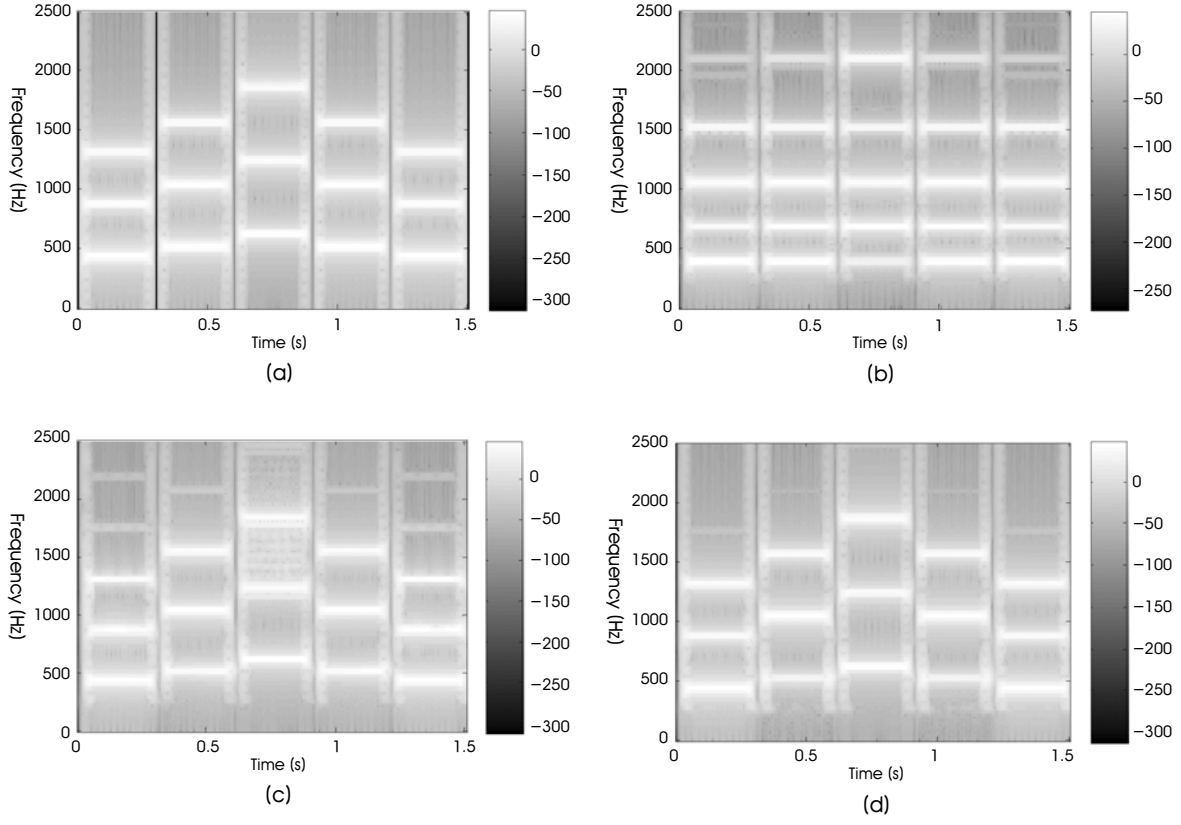

Fig. 4. (a) The spectrogram of a rising-falling melodic contour (the 2nd example in Figure 2 (a)) with three semitones between two successive notes. (b) The spectrograms of the output from the 8 channel conventional speech processing strategy. (c) The spectrograms of the output from the proposed speech processing strategies employing the TEO-based IF encoding. (d) The spectrogram of the output from the proposed speech processing strategies employing the Hilbert-transform-based IF encoding.

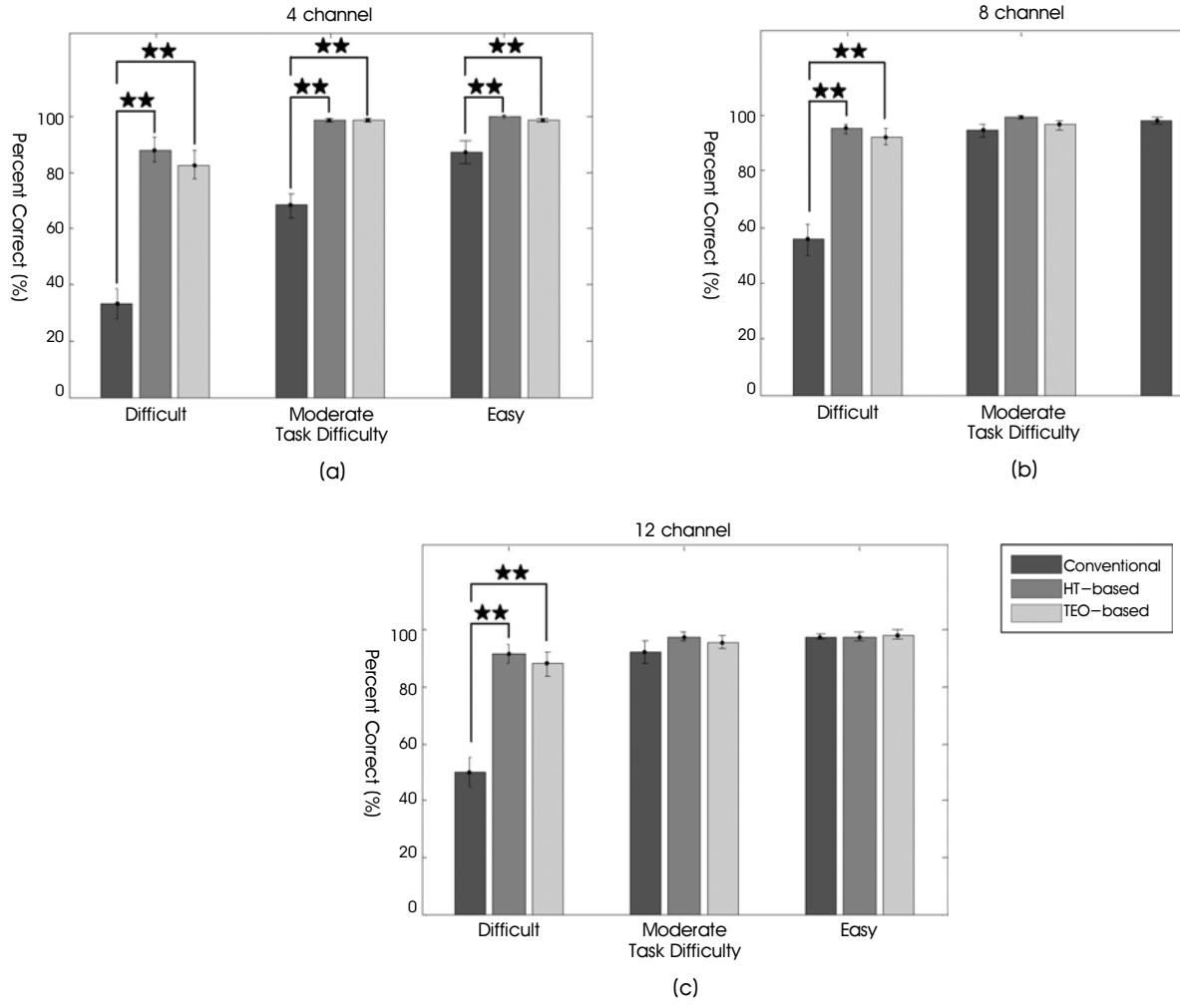

Fig. 5. The melody recognition performances of conventional strategy and the two strategies employing the IF encoding. (a) 4 channel implementation. (b) 8 channel implementation. (c) 12 channel implementation. (★: $p \leq 0.05$, ★★: $p \leq 0.01$)

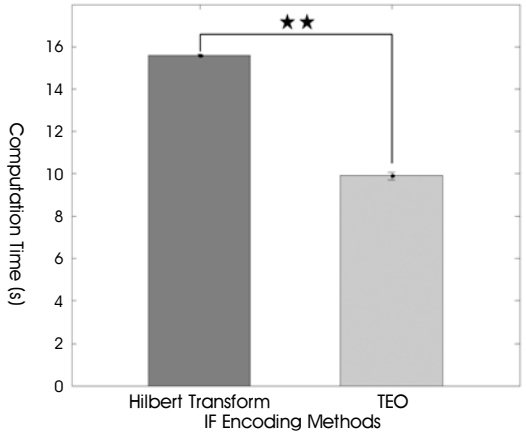

Fig. 6. Thee comparison of the computational times of the Hilbert-transform-based, and the TEO-based strategies (16 channel implementation, ★★: $p \leq 0.01$)).

difference in computation time was statistically significant (t-test, $p<10^{-60}$).

Ⅳ. DISCUSSION

In this paper, we proposed an improved speech processing strategy using the TEO-based IF encoding. The most significant finding is quantitative assessment of melody recognition performances of two strategies based on two specific implementation of IF encoding. The TEO-based and the HT-based methods showed similar IF extraction performances, however, the former required much lower computational cost than the latter. Considering that these two methods showed similar melody recognition performances, the proposed TEO-based IF encoding is judged to be much

better than the HT-based method to be used in CI speech processor, which requires real-time operation with low power consumption for longer battery time.

Although relatively less attention has been paid to the music perception, it is important for CI recipients' quality of life. Satisfactory melody recognition is essential for the music perception. Moreover, for more complete recovery of hearing capability, such as the perception of various environmental sounds, it is meaningful to enable the melody perception of CI recipients. Typically, current CI devices provide 16-24 electrodes at maximum, but the number of effective channels seems to be much less than that of physical channels. This may not cause serious problem for the speech perception under quiet environment. However, for the melody perception, the number of effective channels should be increased or other method is necessary[2y for better representation of the frequency contents. We expect that the IF encoding by the TEO may provide a good alternative to the increment of the number of effective channels for better representation of spectral information.

Conventionally, CI speech processing strategies employ the envelope extraction, and it does not include any dedicated method to encode frequency contents of input other than the subband decomposition by a bandpass filter array. Thus, important frequency-domain information such as pitch and spectral peaks may not be properly represented at the output of the speech processor. This may yield a significant deterioration of melody recognition. As expected, our results showed that the information on the IF made a significant contribution on the melody recognition performance. The two strategies incorporating the IF encoding showed considerable enhancement in the recognition of difficult melodic contours, regardless of the method for extracting the IF information.

The enhanced representation of frequency content may also be helpful for speech recognition. Frequency-domain characteristics of speech waveforms, such as fundamental frequency and formants, are known to be essential for speech recognition under noise. For example, it is recognized that the information on formants is crucial for proper representation of vowels, and also for consonant representation, since formant transition provides valuable information for the identification of consonants such as plosives, stops, and fricatives [16]. It is known that the information on formant frequencies are encoded in the population responses of the auditory nerves [17],[18]. Accurate representation of formant transition may be especially important for Korean, since Korean words include a lot of diphthong vowels [19]. The proposed strategy may also contribute to better recognition of tonal languages such as Chinese [20], since it is expected that the tonal information can be better extracted if the IF is more accurately represented. Further study is planned to investigate the formant representation performance of the proposed strategy, and eventually its speech recognition performance.

The difference in computational loads between the TEO-based and the Hilbert transform-based strategies originate from the computational processes involved. For the Hilbert transform, convolution operation should be performed. Although the fast convolution algorithm based on FFT is usually employed, the TEO-based method involves only a few elementary numerical operations for a few time samples. Thus, the TEO-based speech processor is much more suitable considering real-time processing required for CI devices, than the HT-based strategy. Considering that application-specific integrated circuits (ASICs) or digital signal processors (DSPs) are employed for the implementation of CI speech processors and they should be operated by a battery for a long time, significant reduction in computational load is very helpful. The 35% reduction of computational time was obtained from the Matlab implementation, where a high performance built-in routine for the computation of FFT is used, while the implementation of the TEO is not fully optimized. We expect that the difference in computational load between the two IF encoding methods will be more significant if we compare the comparison is made by optimized implementations in DSPs. However, further elaborated study is necessary to characterize actual benefit of the TEO-based strategy by implementing each strategy under various implementation platforms.

For clinical tests of new strategies, a method for stimulation waveform generation should be devised to properly encode fine structure information. Stimulation waveforms should be determined based on fine structure in addition to envelopes. Several methods has been suggested for this, e.g., by Nie et al. [21] and Throckmortona et al. [22]. They include modulation of pulse rate based on fine structure, simultaneous amplitude-frequency modulation of analog waveforms, and use of multiple carrier frequency for each channel. However, none of them were clinically tested by actual implementation of strategy. The clinical test necessitates alteration embedded program of CI device and long-term training with multiple CI patients, and thus, only one preliminary result on clinical test

of fine structure encoding has been recently reported to our knowledge [23]. Despite this difficulty, further efforts should be continued on clinical tests to assess the benefits obtained by incorporating fine structure foe CI sound processing strategy.

ACKNOWLEDGMENT

This study was financially supported by a grant from the Korea Health 21 R&D Project, Ministry of Health & Welfare (grant no. A050251), and by a grant from the Industrial Source Technology Development Program (no. 10033812) of the Ministry of Knowledge Economy (MKE) of Korea.

REFERENCE

[1] B. Wilson, and C. Finley, "Improved speech recognition with cochlear implants," *Nature*, vol. 352, pp. 236-238, 1991.

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

[3] P. Loizou, M. Dorman, and Z. Tu, "On the number of channels needed to understand speech," *J. Acoust. Soc. AM.*, vol. 106, pp. 2097-2103, 1999.

[4] B.S. Wilson, R. Schatzer, E.A. Lopez-Poveda, X.A. Sum, D.T. Lawson, and R.D. Wolford, "Two new directions in speech processor design for cochlear implants, "*Ear. Hearing*, vol. 26, pp. 73-81, 2005.

[5] Z.M. Smith, B. Delgutte, and A.J. Oxenham, "Chimaeric sounds reveal dichotomies in auditory perception," *Nature*, vol. 416, pp. 87-90, 2002.

[6] S. Brill, A. Moltner, W. Harnisch, J Muller ,and R. Hagen, "Temporal fine structure coding in low frequency channels: speech and prosody understanding, pitch and music perception and subjective benefit evaluated in a prospective randomized study," presented at 2007 Conference on Implantable Auditory Prostheses, Lake Tahoe, California, 2007.

[7] W.R. Drennan, J.K. Longnion, C. Ruffin, and J.T. Rubinstein, "Discrimination of Schroeder-phase harmonic complexes by normal-hearing and cochlear-implant listeners," *J. Assoc. Res. Otolaryngol*., vol. 9, pp. 138-149, 2008.

[8] W.R. Drennan, J.H. Won, V.K. Dasika, and J.T. Rubinstein, "Effects of temporal fine structure on the lateralization of speech and on speech understanding in noise," *J. Assoc. Res. Otolaryngol*., vol. 8, pp. 373-383, 2007.

[9] F. Zeng, K. Nie, G.S. Stickney, Y. Kong, M. Vongphoe, A. Bhargave, C. Wei, and K. Cao, "Speech recognition with amplitude and frequency modulations," *Proc. Natl. Acad. Sci. USA*, vol. 102, pp. 2293-2298, 2004.

[10] K. Nie, G. Sticney, and F. Zeng, "Encoding frequency modulation to improve cochlear implant performance in noise," *IEEE Trans. Biomed. Eng*., vol. 52, pp. 64-73, 2004.

[11] R.E. Ziemer, and W.H. Tranter, Principles of Communications, 5th Ed., New York: Wiley, 2001.

[12] J.F. Kaiser, "On a simple algorithm to calculate the energy of a signal, "*in Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, vol. 1, pp. 381-384, 1990.

[13] P. Maragos, J.F. Kaiser, and T.F. Quatieri, "Energy separation in signal modulations with application to speech analysis," *IEEE Trans. Signal Process*., vol. 41, pp. 3024-3050, 1993.

[14] A. Potamianos, and P. Maragos, "A comparison of the energy operator and the Hilbert transform approach to signal and speech demodulation," *Signal Process*., vol. 37, pp. 95-120, 1994.

[15] J.H. Galvin, Q. Fu, and G. Nogaki, "Melodic contour identification by cochlear implant listeners," *Ear and Hear*., vol. 28, pp. 302-310, 2007.

[16] S. Bandyopadhyay, E.D. Young, "Discrimination of voiced stop consonants based on auditory nerve discharges," *J. Neurosci*., vol. 24, pp. 531-541, Jan. 2004.

[17] E.D. Young, M.B. Sachs, "Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers," *J. Acoust. Soc. Am*., vol. 66, pp. 1381-1403, Nov. 1979.

[18] A.R. Palmer, I.M. Winter, C.J. Darwin, "The representation of steady-state vowel sounds in the temporal discharge patterns of the guinea pig cochlear nerve and primary-like cochlear nucleus neurons," *J. Acoust. Soc. Am.*, vol. 79, pp. 100-113, Jan. 1986.

[19] B.K. Yang, "An Acoustical Study of Korean Diphthongs," *MALSORI (in Korean)*, vol. 25, pp. 3-26, 1993.

[20] Y. Kong, and F. Zeng, "Temporal and spectral cues in Mandarin tone recognition," *J. Acoust. Soc. AM*., vol. 120, pp. 2830-2840, 2006.

[21] K. Nie, G.S. Stickney, and F.G. Zeng, "Encoding frequency modulation to improve cochlear implant performance in noise," *IEEE Trans. on Biomedical Engineering*, vol. 52, pp. 64-73, 2005.

[22] C.S. Throckmortona, M. S. Kucukoglua, J. J. Remusa, L. M. Collins, "Acoustic model investigation of a multiple carrier frequency algorithm for encoding fine frequency structure: Implications for cochlear implants," *Hearing Research*, vol. 218, Issues 1-2, pp. 30-42, 2006.

[23] R Schatzer, A Krenmayr, D. K. Au, M. Kals, C. Zierhofer, "Temporal fine structure in cochlear implants: preliminary speech perception results in Cantonese-speaking implant users," Acta Otolaryngol., vol. 130, no. 9, pp. 1031-1039, 2010.