Bionic Wavelet Transform: A New Time–Frequency Method Based on an Auditory Model

Jun Yao, *Student Member, IEEE* and Yuan-Ting Zhang*, *Senior Member, IEEE*

***Abstract*—In this paper, a new adaptive wavelet transform, named bionic wavelet transform (BWT), is developed based on a model of the active auditory system. The most distinguishing characteristic of BWT is that its resolution in the time–frequency domain can be adaptively adjusted not only by the signal frequency but also by the signal instantaneous amplitude and its first-order differential. The automatically adjusted resolution, even in a fixed frequency along the time-axis, is achieved by introducing the active control of the auditory system into the wavelet transform (WT). Other properties of BWT include that: 1) BWT is a nonlinear transform that has high sensitivity and frequency selectivity; 2) BWT represents the signal with a concentrated energy distribution; and 3) the inverse BWT can reconstruct the original signal from its time–frequency representation. In order to compare these three properties between BWT and WT, experiments were conducted on both constructed signals and real speech signals. The results show that BWT performs better than WT in these three aspects, and that BWT is appropriate for speech signal processing, especially for cochlear implants.**

***Index Terms*—Bionic wavelet transform, cochlear implants, otoacoustic emissions, time–frequency analysis.**

I. Introduction

BASED on an active auditory model, a new time–frequency method, named bionic wavelet transform (BWT), is developed in this paper. BWT extends one-dimension-adjustable-resolution (1D-adjustable-resolution) along the frequency-axis (f-axis) in traditional wavelet transform (WT) to two-dimension-adjustable-resolution (2D-adjustable-resolution) along both the time axis (t-axis) and the f-axis. By introducing the active mechanism of a bio-system into WT, good tradeoff between time and frequency resolutions can be achieved. The consequent transform, BWT, has properties that are appropriate for speech signal processing, especially for cochlear implants.

Extending 1D-adjustable-resolution to 2D-adjustable-resolution for better tradeoff between time and frequency resolutions has been a topic in WT for a long time. It is well known that WT allows 1D-adjustable-resolution along the f-axis, while in a fixed frequency along the t-axis, WT is the same, in terms of resolution, as short-time Fourier transform (STFT). Research has proven that the choice of mother wavelet function dramatically affects the appearance and quality of the resultant time–frequency representation [1], which reflects the importance of the 2D-adjustable-resolution problem. To solve this problem, many attempts have been made to apply different mother functions at different times and frequencies. Among them, the state-of-the-art methods based on WT include adaptive WT (AWT) and wavelet packets (WP), both of which select mother functions according to an entropy criterion. Some papers have shown that AWT or WP achieves good tradeoff between resolutions and, thus, has good applications in data compression and representation [2], [3]. Applications of AWT and WP in bio-signal processing, however, are limited by the facts that no entropy criterion is known to be appropriate for biomedical applications and that the computation cost of AWT and WP is usually very high. Therefore, further studies on the 2D-adjustable-resolution problem for biomedical applications are worthwhile.

The hypothesis behind developing BWT for biomedical applications is that a proper introduction of a bio-model into a signal processing method benefits the application of this method to corresponding bio-signal processing. Supportive evidence is that the development of signal processing methods for bio-signal processing forms a trend of mimicking the bio-system. Without a doubt, Fourier transform is a popular and helpful method of representing signals, including complex biomedical signals in the frequency domain. And its derivative, STFT, has been found to be very similar to many signal processing mechanisms in bio-systems. For example, Gabor Transform, a well-defined STFT, gives a mathematical abstract of the signal processing mechanism of the human vision field both in the function itself and in the bandpass filter (BPF) bank property [4]. However, in different center frequencies, the BPFs in STFT have different quality factors varying within a range far beyond a reasonable value for human vision and auditory systems. In fact, in many bio-systems, the quality factors vary in a small range, which for the human cochlea is from 5 to 15 [5]. Usually, the filters in human auditory and visual systems are viewed as constant-Q filters. These facts explain why WT, which has a constant Q, is viewed to work in a mechanism more similar to bio-systems than STFT. It seems that the progress in signal processing tends to be more similar to working mechanisms of bio-systems. It is worthwhile exploring the extension of WT with 1D-adjustable-resolution to a new method with 2D-adjustable-resolution, inspired by mechanisms of bio-systems.

In this study, we attempted to introduce the active mechanism of the auditory system into WT. Our method is supported by the existence of 2D-adjustable-resolution in the auditory system and by the fact that this resolution is necessary to

Manuscript received September 4, 2000; revised April 27, 2001. *Asterisk indicates corresponding author.*

J. Yao is with The Chinese University of Hong Kong, Shatin, N. T., Hong Kong.

*Y.-T. Zhang is with The Chinese University of Hong Kong, Shatin, N. T., Hong Kong (e-mail: ytzhang@ee.cuhk.edu.hk).

Publisher Item Identifier S 0018-9294(01)06154-7.

maintain normal hearing. Early experiments on the cochlear partition in human and animal cadavers found that the basilar membrane in the cochlea had frequency selectivity [6], which showed that a passive cochlea analyzes the signal in the time–frequency domain. From the point of view of signal processing, passive cochlear models are linear BPF banks and work like WT. The findings of otoacoustic emissions (OAEs), which are the acoustic energy generated in the cochlea and can be recorded at the ear canal [7], after the 1970s provided evidences that the cochlea is an active nonlinear system. It is widely accepted today that the generation of OAEs is related to the active mechanism of the cochlea *in vivo*, and that this active mechanism is associated with activities of outer hair cells (OHCs) [8]. One important role of OHCs is that they provide 2D-adjustable-resolution and signal amplification, which is necessary to keep the high sensitivity and frequency selectivity of the cochlea *in vivo* [9]. Damage to OHCs caused by high noise, drugs, or other factors may lead to impairment or even total loss of hearing [8].

The importance of the active mechanism in the auditory system motivates us to introduce its model into WT. To do this, i.e., to extend WT to BWT, a model of the active mechanism and a way of introducing this model to WT are required. The cochlear model, which has distinct active and passive mechanisms, has not been quantitatively established because of the difficulty in directly detecting the active mechanism from the cochlea. Fortunately, based on OAEs, we can develop a qualitative cochlear model. Based on this model, developing a new time–frequency method is, therefore, possible. The way to introduce the model into WT is developed based on the analysis of the similarity and difference between the biological model and WT.

To show that BWT has 2D-adjustable-resolution and the following new properties: 1) better tradeoff between time and frequency resolutions than WT; 2) concentrated signal energy distribution in BWT representation; and 3) existence of inverse transform, we carried out computer simulations on constructed signals and experiments on real speech signals. To show that BWT is appropriate for speech signal processing for biomedical applications, we gave a simple example of the BWT application to speech signal processing in cochlear implants. Our results support our argument that 2D-adjustable-resolution can be achieved by BWT, which has properties that are appropriate for speech signal processing, especially for cochlear implants.

II. The Ear Model Based on Otoacoustic Emissions

In previous work, a nonlinear auditory model by Giguere was adopted and modified as an OAE model [10]–[12]. In this model, a point of the basilar membrane (BM) is modeled by the following equation:

$$\ddot{d}(x,t) + R_{eq}(x,d)\dot{d}(x,t)/L(x) + \omega_0^2(x)d(x,t) = P \quad (1)$$

where

x distance along the BM from the basal end;

t time;

d displacement of the BM;

$\dot{d}$ and $\ddot{d}$ first- and second-order differentials of d in terms of t;

P pressure difference across the BM;

ω_0 character frequency of the point and equals $1/\sqrt{L(x)C(x)}$, where $L(x)$ and $C(x)$ represent the acoustic mass and compliance, respectively.

The equivalent resistance, R_{eq}, in (1) is given by

$$R_{eq} = R(x) - G_1(x)\frac{d_{1/2}}{d_{1/2} + |d(x,t)|}R(x) \quad (2)$$

where

$R(x)$ passive resistance corresponding to the acoustic resistance;

$d_{1/2}$ saturation factor;

$G_1(x)$ active gain factor whose value is related to the activity of the corresponding OHC-group.

The second term in (2) represents the active resistance function of the OHC control.

In Giguere's model, the analogy between acoustics and electronics determines passive parameters, among which $L(x)$ is determined by the acoustic mass; $\omega_0(x)$ is determined by the cochlear frequency-position function [13]; $C(x)$ is determined by $1/\omega_0^2 L(x)$; and $R(x)$ is obtained by $Q^{-1}\sqrt{L(x)/C(x)}$, where $Q = 2$ [14]. All these parameters, R, L, C, Q, and ω_0, are constants for a fixed x.

Clearly from (1), a point of BM is modeled by a BPF that has a nonlinear damping. This model had been proven to be useful for the description of some major features of OAEs [10]–[12]. Recent research results, however, have shown that nonlinear damping alone is not enough to describe the active mechanism of the cochlea, and that nonlinear compliance is also necessary [15]. We further modified this model by introducing the nonlinear capacitance

$$C_{eq}(x) = \left(1 + G_2(x)\left|\frac{\partial[d(x,t)]}{\partial t}\right|\right)^2 C(x) \quad (3)$$

where $G_2(x)$ is the active factor that associates with the effect of the active mechanism on the compliance on a point of BM. The modified model has the same formula as (1), but ω_0 equals $1/\sqrt{L(x)C_{eq}(x)}$. Simulation results show that some improvements in the description of OAE properties can be achieved by this modification. The modified auditory model can describe the property of an active cochlea.

The purpose of using this model in this study is not to detect the active mechanism of the auditory system perfectly, but to show the possible effect of an active mechanism on the signal processing.

III. Definition of Bionic Wavelet Transform

A. Introduction of the Active Mechanism into Traditional Wavelet Transform

To introduce the active mechanism of the auditory system into WT, we replaced the constant Q_0 of WT by a variable Q_T. In the active model, without the active mechanism (i.e., $G_1(x) = G_2(x) = 0$), the passive model of a point of BM has a constant quality factor $Q = R^{-1}\sqrt{L/C}$. With the introduction of the active mechanism, however, the equivalent quality factor, $Q_{eq} = R_{eq}^{-1}\sqrt{L/C_{eq}}$, varies with the displacement of BM. We determined the "transfer" function (T-function), which transfers the constant quality factor Q of the passive model to the variable quality factor Q_{eq} of the active model. Using the same T-function, we transferred the constant quality factor Q_0 of WT to the variable quality factor Q_T of BWT. We named it bionic WT because the T-function stems from the bio-system model.

B. Definition of Bionic Wavelet Transform

A mother function, $h(t)$, of WT must satisfy the admissible condition, which implies that $h(t)$ has some oscillations. If the window center of $h(t)$ is at f_0 in the frequency domain, we can represent $h(t)$ as its envelope function, $\tilde{h}(t)$, modulated by the sinusoidal signal at frequency f_0

$$h(t) = \frac{1}{\sqrt{a}}\tilde{h}(t)\exp(j\omega_0 t) \tag{4}$$

where $\omega_0 = 2\pi f_0$. If the signal to be analyzed is $f(t)$, the WT is defined as

$$(WTf)(\tau, a) = \frac{1}{\sqrt{a}}\int f(t)\tilde{h}^*\left(\frac{t-\tau}{a}\right) \cdot \exp\left(-j\omega_0\left(\frac{t-\tau}{a}\right)\right) dt \tag{5}$$

where

a	scale;
τ	time shift;
* and $\int\cdot$	complex conjugate and the integration from negative infinity to positive infinity, respectively.

To mimic the OHC-like control, a new parameter T, where $T > 0$, is introduced into the WT mother function resulting in the BWT mother function

$$h_T(t) = \frac{1}{T\sqrt{a}}\tilde{h}\left(\frac{t}{T}\right)\exp(j\omega_0 t). \tag{6}$$

With the new mother function, we define BWT as

$$(BWT_T f)(\tau, a) = \frac{1}{T\sqrt{a}}\int f(t)\tilde{h}^*\left(\frac{t-\tau}{aT}\right) \cdot \exp\left(-j\omega_0\left(\frac{t-\tau}{a}\right)\right) dt. \tag{7}$$

Clearly from (7), the centers of the analyzing window in the time and frequency domains are the same as those of WT. The only difference is that the envelope of the BWT mother function can be adjusted by the parameter T. The introduction of the factor $1/T$ in (6) is for normalization.

Given T as a constant in a short enough period, the Fourier transforms of $h(t)$ and $h_T(t)$ are

$$H(\omega) = \frac{1}{\sqrt{a}}\tilde{H}(\omega - \omega_0) \tag{8}$$

and

$$H_T(\omega) = \frac{1}{\sqrt{a}}\tilde{H}[T(\omega - \omega_0)] \tag{9}$$

respectively, where $\tilde{H}(\omega)$ is the Fourier transform of $\tilde{h}(t)$. From (8) and (9), it is clear that the quality factor Q_T of $h_T(t)$ for BWT is related to Q_0 of $h(t)$ for WT by

$$Q_T = TQ_0. \tag{10}$$

The T-function in (10) is introduced from the active auditory model. In the model, when substituting (2) and (3) to $Q_{eq} = R_{eq}^{-1}\sqrt{L/C_{eq}}$, Q_{eq} is related to the constant Q by

$$Q_{eq} = \left(1 - G_1\frac{d_{1/2}}{d_{1/2} + |d|}\right)^{-1}(1 + G_2|\partial d/\partial t|)^{-1}Q. \tag{11}$$

Comparing (11) with (10), the T-function is given by

$$T(\tau + \Delta\tau) = \left(1 - \tilde{G}_1\frac{BWT_s}{BWT_s + |BWT_f(\tau, a)|}\right)^{-1} \times \left(1 + \tilde{G}_2|\partial(BWT_f(\tau, a)/\partial(t)|\right)^{-1} \tag{12}$$

where

$BWT_f(\tau, a)$	coefficient of the BWT at time τ and scale a, which maps the output of the BM filter, d;
BWT_s	saturation constant that maps $d_{1/2}$ in the auditory model;
$\tilde{G}_1$ and $\tilde{G}_2$	active factors;
$\Delta\tau$	calculation step.

By increasing $\tilde{G}_1$ and $\tilde{G}_2$, resolutions in frequency domain and time domain can be increased, respectively.

To obtain (9), T must be a constant in the period of $\Delta\tau$. This approximation is reasonable if the signal and its first differential are continuous and $\Delta\tau$ is small enough. Clearly from the definition, BWT is a nonlinear transform.

IV. New Properties of Bionic Wavelet Transform

To show the new properties of BWT, computer simulations and experiments were conducted on constructed signals and real speech signals, respectively. When doing these experiments, we set the mother function for both WT and BWT as Morlet function to achieve the smallest product of window size in time and frequency domains. The parameters of Morlet function were chosen according to the auditory model presented in Section II, where the center frequency of the mth scale is $f_m = f_0/q^m$, where f_0 =15 165.4 Hz, q =1.035 295 2, and $m = 0, 1, 2, \ldots$, 127 [11]. The constant $T_0 = 0.0005$ was used in the WT, and a dynamic adjusted T according to (12), where $\tilde{G}_1 = 0.87$, $\tilde{G}_2 = 45$, and $BWT_s = 0.8$, was used in the BWT. The choice of values for $\tilde{G}_1$, $\tilde{G}_2$, and BWT_s depends on the signal properties and other parameter settings. These parameters are practical in the numerical computations.

A. Resolutions in Time and Frequency Domains of BWT

The radii of an analysis window in time and frequency domains were used to evaluate the resolutions. If the radii of the window for WT mother function $h(t)$ in time and frequency domains are represented by Δ and $\hat{\Delta}$, respectively, then the analyzing window of WT at a certain time $t_0$and scale a is

$$[t_0 - a\Delta, t_0 + a\Delta] \times \left[\frac{\omega_0}{a} - \frac{\hat{\Delta}}{a}, \frac{\omega_0}{a} + \frac{\hat{\Delta}}{a}\right] \tag{13}$$

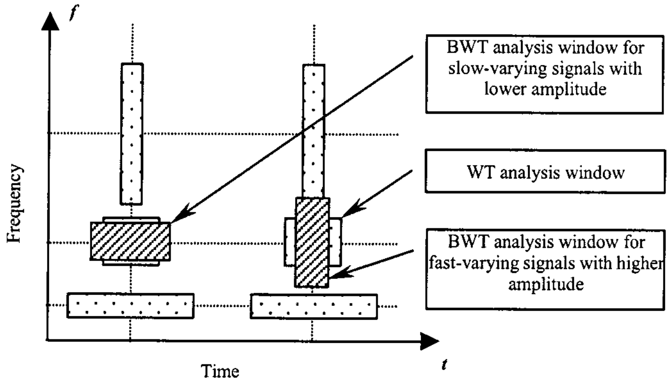

Fig. 1. The resolutions in the time and frequency domains for BWT and WT. The window size of the WT (see the dotted windows) varies with the change of analyzing frequency. In the WT, for a fixed mother function, all the windows in a certain scale along the t-axis are fixed. However, both the time and frequency resolutions can be different in the BWT even in a certain scale (see the shadowed windows). The adjustment of the BWT resolution in the same scale is controlled by T-function, which is related to the signal instantaneous amplitude and its first-order differential.

while that of BWT with the mother function $h_T(t)$ is

$$[t_0 - aT\Delta,\, t_0 + aT\Delta] \times \left[\frac{\omega_0}{a} - \frac{\hat{\Delta}}{aT},\, \frac{\omega_0}{a} + \frac{\hat{\Delta}}{aT}\right]. \tag{14}$$

Clearly from (13), the radii of the WT window in time and frequency domains are $a\Delta$ and $\hat{\Delta}/a$, respectively, which are not related to the local time-varying characteristics of the signal. On the contrary, the radii of the BWT window in time and frequency domains are $aT\Delta$ and $\hat{\Delta}/(aT)$, respectively, where T is related to signal instantaneous amplitude and its first-order differential. Clearly from (12) and (14), BWT achieves 2D-adjustable-resolution with the characteristics that higher frequency resolution and lower time resolution are used to decompose a slow-varying signal with lower amplitude, and that for a fast-varying signal with higher amplitude, lower frequency resolution and higher time resolution are used. The relationship between resolutions and signal amplitude is consistent with the phenomena that when we talk in a lower voice, we usually talk much more slowly to achieve a better degree of hearing.

The difference in resolutions between WT and BWT is further illustrated in Fig. 1, where all analysis windows of WT, presented by dotted windows, are the same if they are on the same scale. However, analysis windows of BWT, presented by shadowed windows, vary with the local time-varying feature of the signal, even on a fixed scale.

To show that BWT results in better tradeoff between time and frequency resolutions than WT, a constructed signal given by

$$f(t) = \begin{cases} \sin(2\pi \times 2000 \times t), & 0 \le t \le 0.01 \text{ s} \\ \sin(2\pi \times 2500 \times t), & 0.01 \text{ s} < t < 1 \text{ s} \end{cases} \tag{15}$$

was used. The signal was sampled with frequency at 50 kHz. Because the signal has two different frequency components, which change sharply from 2 to 2.5 kHz at time 0.01 s, the contour plot of the ideal time–frequency representation of this signal should have a line spectrum at the two given frequencies, and a sharp frequency transient at time 0.01 s. We can clearly see from the contour plots of the signal time–frequency representations by BWT and WT in Fig. 2(a) and (b) that BWT achieves better tradeoff between time and frequency resolutions than WT. Although 128 scales were used in our analysis, only the scales from 50 to 63 are plotted. The corresponding center frequencies of these scales are listed in Table I.

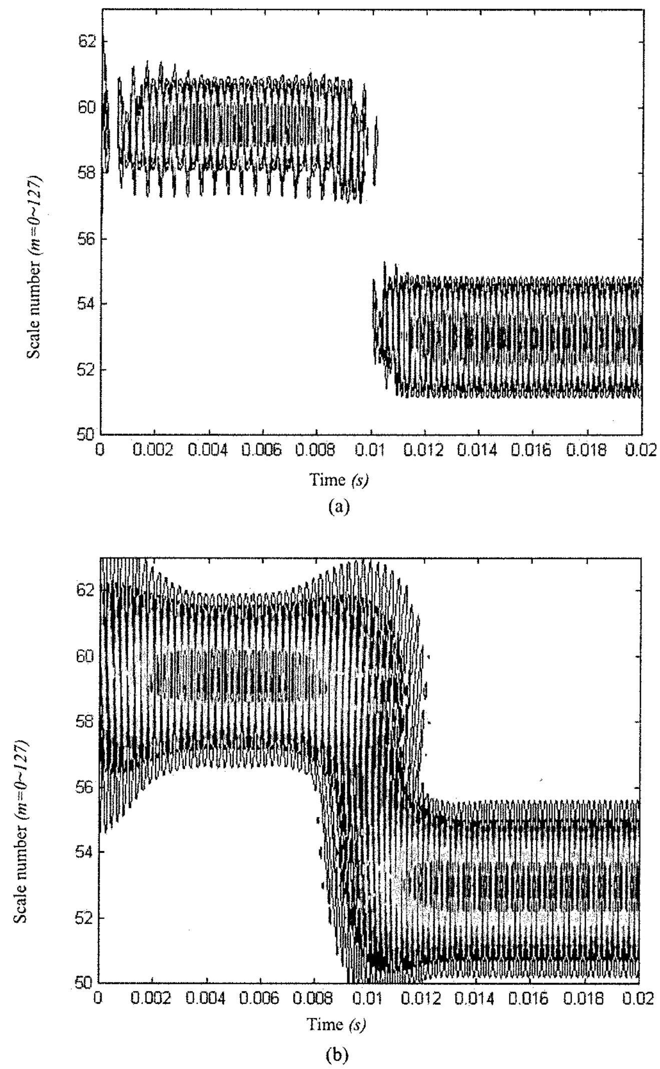

Fig. 2. Contour plot of signals decomposed by (a) BWT and (b) WT. The center frequency of the mth scale is $f_m = f_0/q^m$, $m = 0 \sim 127$, $f_0 = 15\,165.4$ Hz, and $q = 1.035\,295\,2$.

TABLE I
THE CENTER FREQUENCIES FOR SCALES FROM 50 TO 63

Scale number	Center frequency (Hz)
50	2771.456
51	2676.972
52	2585.709
53	2497.557
54	2412.411
55	2330.167
56	2250.727
57	2173.996
58	2099.88
59	2028.291
60	1959.143
61	1892.352
62	1827.838
63	1765.523

B. Concentrated Energy Distribution of BWT

One of the goals of performing a transformation is to achieve a more concentrated distribution of energy in the new space supported by the kernel function. In this experiment, four real speech signals, "bat," "bit," "put," and "pig," spoken by a single female were used to compare the energy distribution of BWT and WT. A personal computer with a Crystal PnP sound card recorded these speech signals at a sampling frequency of 22 050 Hz and 16 bits. Then signals were down-sampled to a sampling frequency of 11 025 Hz, and then processed by WT or BWT with 128 scales in total.

When conducting the experiment, we first normalized the decomposed signal in the time–frequency domain. Then we chose a threshold and set the coefficients, whose absolute values were less than the threshold, at zero, and retained the other ones. For a discrete signal f_n, $n = 0, 1, 2, \ldots, N-1$, after being decomposed by WT or BWT with M ($M = 128$ in this experiment) scales and with $\Delta\tau = 1/f_s$, where f_s is the sampling frequency, the total number of decomposed coefficients is $N \times M$, and the energy of the decomposed signal is $E = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} Coef_f^2(m, n)$, where $Coef_f(m, n)$ is the coefficient of BWT or WT at time $n \times \Delta\tau$ and scale q^m. Since only the coefficients, whose absolute values were no less than the threshold, were retained, some of the signal energy is lost. We suppose that if the energy of the signal has a concentrated distribution in the space supported by a kernel function, more signal energy can be retained with a smaller number of coefficients. After selecting the coefficients, we calculated the ratio of the retained-coefficient number, which is defined as (the number of retained coefficients)/($N \times M$), and the ratio of retained energy, which is defined as (the energy of retained coefficients)/E. The results obtained with different thresholds are plotted in Fig. 3, where the solid lines and the broken lines are the results of BWT and WT, respectively. As can be seen in Fig. 3, the results obtained for all tested signals show that BWT always has more retained energy with the same number of retained coefficients than WT, or always requires a smaller number of coefficients for the same retained energy. This result implies that BWT represents the signal with a concentrated energy distribution.

C. BWT Inverse Transform

The inverse BWT (IBWT) is defined as follows:

$$f(t) = \frac{1}{C_{h_T}} \iint BWT_f(a, \tau) h_T\left(\frac{t-\tau}{a}\right) d\tau \frac{da}{a^2}. \quad (16)$$

where $C_{h_T} = \int \frac{|\tilde{H}[T(\omega-\omega_0)]|^2}{|\omega|} d\omega$. We can prove that if the signal and its first-order differential are continuous, BWT can reconstruct the original signal without distortion [16]. This is also supported by an experimental result conducted on the real speech signal "put," which is one of the speech signals used in the experiment of Section IV-B. Fig. 4 (a)–(c) shows the original signal and reconstructed signals by IBWT and inverse WT (IWT), respectively. When doing the inverse transform, we first normalized the signal in the time–frequency domain, and then only retained the coefficients whose absolute values were larger than 0.003, at which level about 98% of the energy was retained for both IBWT and IWT. Coefficients less than 0.003 were set at zero. Correlation coefficients between the original signal and the reconstructed signal are 0.7783 for BWT and 0.6875 for WT, which shows that the quality of the reconstructed signal by IBWT has, at least, a comparable quality to that of the reconstructed signal by IWT. This result supports that if a proper T-function and calculation parameters are used, IBWT can reconstruct the signal. Constraints on the T-function for reconstructing signals are derived in [16].

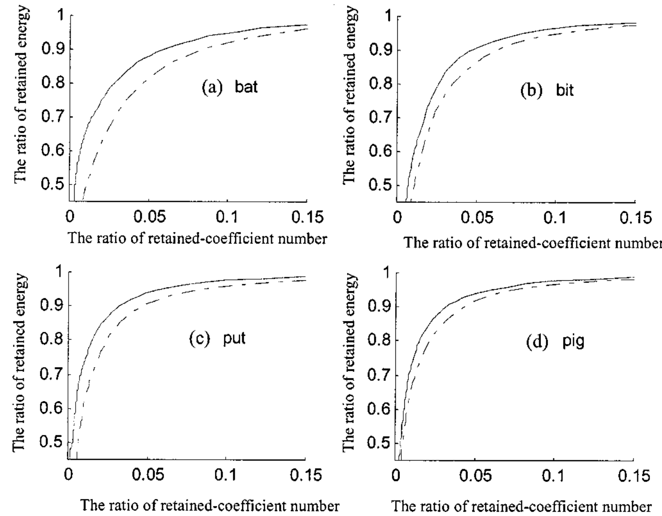

Fig. 3. The ratio of retained energy versus of the ratio of retained-coefficient number of BWT (solid lines) and WT (broken lines) for real speech signals (a) "bat," (b) "bit," (c) "put," and (d) "pig." The ratio of retained-coefficient number is defined as (the number of retained coefficients)/($N \times M$) and the ratio of retained energy is defined as (the energy of retained coefficients)/E, where N is the data length in time domain; M is the total number of scales; E is the energy of the decomposed signal.

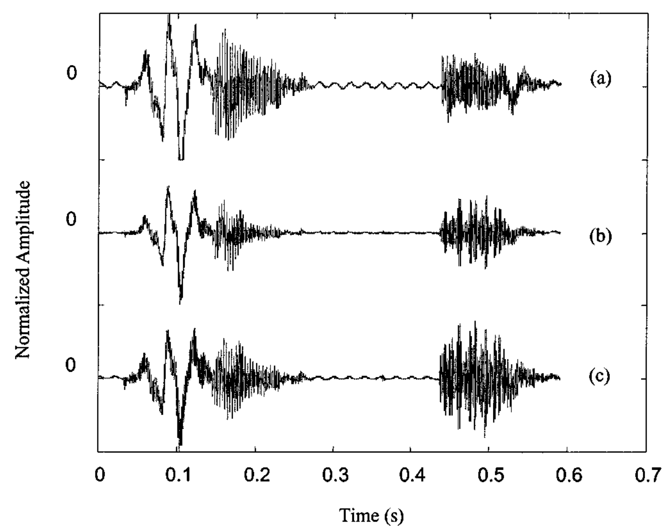

Fig. 4. The results of reconstructed signals by BWT and WT: (a) the original signal and reconstructed signals by the inverse transform of (b) BWT and (c) WT.

V. Example of BWT Application to Speech Signal Processing in Cochlear Implants

To show BWT is appropriate for speech signal processing, an example of BWT application given in this paper is cochlear implants (CI). The purpose of CI is to use an electrode or an array of electrodes to stimulate the hearing nerve directly. CI successfully restores partial hearing to people whose auditory sensors (the hair cells in the cochlea) are not functional. The state-of-the-art signal processing strategies used in CI are based on BPFs, which typically use separate hardware to implement filtering, noise removal, spectral estimation, and equalization. With these strategies, the CI products are difficult to adjust to fit the individual CI subject because too many stages are involved in implementation [17]. New strategies based on WT were proposed because it was possible to integrate the various signal-processing requirements into a single digital signal processor [17], [18]. Neither the strategies based on BPFs nor the ones based on WT are as effective as a healthy ear because the auditory system is a nonlinear system providing high sensitivity and frequency selectivity to analyze the signal. BWT stems from the active auditory model and is a nonlinear transform so has higher sensitivity and frequency selectivity than WT. Therefore, we attempted to test the application of BWT to speech signal processing for CI application.

Preliminary tests of BWT application to speech signal processing in CI were conducted on five consonants, /t k d p g/, in the context /iCi/, where "C" represents the consonant. The material is a subset of the Iowa consonant test for CI, and was spoken by a single female speaker in an environment without soundproof. Speech signals were recorded and sampled with the same equipment and the same configurations as used in Section IV. The total number of scales for both BWT and WT was 22, i.e., the speech signal was decomposed to 22 different channels with different center frequencies ranging from 225 Hz to 5291 Hz. The mother function for both BWT and WT was still Morlet function. After decomposing, every 4 ms, signals in the four channels that had the maximum energy were selected. Finally, the speech signal was reconstructed by the inverse transforms from the retained coefficients in the time–frequency domain. According to this experimental method, the number of retained coefficients for both BWT and WT is the same for a specific speech signal.

The ratios of retained energy, root-mean-square (rms) errors for the ith consonant [defined as

$$E_i = \sqrt{\frac{\sum_{n=0}^{N-1}(\overline{f}_{\mathrm{in}} - f_{\mathrm{in}})^2}{N}}$$

where $\overline{f}_{\mathrm{in}}$ is the reconstructed signal and f_{in} is the original signal], and correlation coefficients between the reconstructed signal and the original signal are plotted in Fig. 5 (a)–(c), respectively, where solid and open circles, squares, and triangles represent, respectively, the ratios of retained energy, rms errors, and correlation coefficients for BWT and WT. The average rms errors for BWT and WT are 0.0734 and 0.0946, and the average correlation coefficients for BWT and WT are 0.9337 and 0.8928. The sound quality of the reconstructed signal, judged by a panel of 15 normal hearing subjects, shows that the quality of the reconstructed signal of BWT is significantly improved compared with the quality of the reconstructed signal of WT ($t = 9 > 4.073$, $p = 0.001$) [19].

VI. Discussion and Conclusion

In this paper, a new time–frequency analysis method, BWT, has been developed by introducing the active mechanism of the auditory system into WT. BWT provides new and effective properties for speech signal processing. The motivation to develop BWT is our belief that a signal processing method based on a bio-system is appropriate for bio-signal processing.

Simulation results based on the constructed signal in Section IV-A support that the BWT based on the auditory model achieves 2D-adjustable-resolution, and that BWT has high sensitivity and frequency selectivity. Experimental results based on real speech signals in Section IV-B show that BWT represents the signal with a concentrated energy distribution, which implies that with the same data, more signal energy can be retained. Furthermore, results in Section IV-C based on the speech signal "put" show that under proper conditions, IBWT can reconstruct the signal, thus, we can expect that a better quality of sound can be achieved when more signal energy is retained. This is further supported by the experimental results on BWT and WT applications to consonant processing represented in Section V. Fig. 5(a) shows that more energy is retained by BWT than that by WT. And, thus, from Fig. 5(b) and (c), it is clear that lower rms errors and higher correlation coefficients were obtained by BWT than those by WT. Therefore, BWT properties of concentrated energy distribution and the existence of IBWT are appropriate for biomedical applications, such as biomedical signal encoding and decoding. A simple example of BWT application to speech signal processing in CI supports that BWT is appropriate for cochlear implant applications.

From the point of view of signal processing, BWT can be viewed as a special subset of AWT), because BWT also uses the prior information of the signal to select the proper wavelet function. The distinguishing characteristic of BWT is that it does not use an entropy function as a criterion, but introduces the active control mechanism in the auditory system to adjust the wavelet function. One reason for carrying out wavelet function adjustment according to the active control mechanism in a bio-system is that it is hard to know what entropy function the bio-system uses. The active mechanism, however, is more and more clearly based on the bio-system model and physiological experiments.

Just as the performance of AWT is highly dependent on the entropy function, the performance of BWT is highly dependent on the T-function. In this paper, the T-function stemming from the auditory model performs well when being applied to speech signal processing. We believe that with the improvement of the auditory model, a better performance of BWT can be achieved.

Generally, for different applications, different T-functions should be used. A possible choice of T-function can be made according to the bio-signal under study. For example, for the speech signal, the T-function based on the auditory system is a natural choice. While for the image processing, an appropriate choice of T-function may be the control function in the vision system.

In summary, the active mechanism of the auditory model is introduced into the traditional WT, which leads to the BWT. The special properties of BWT are summarized below:

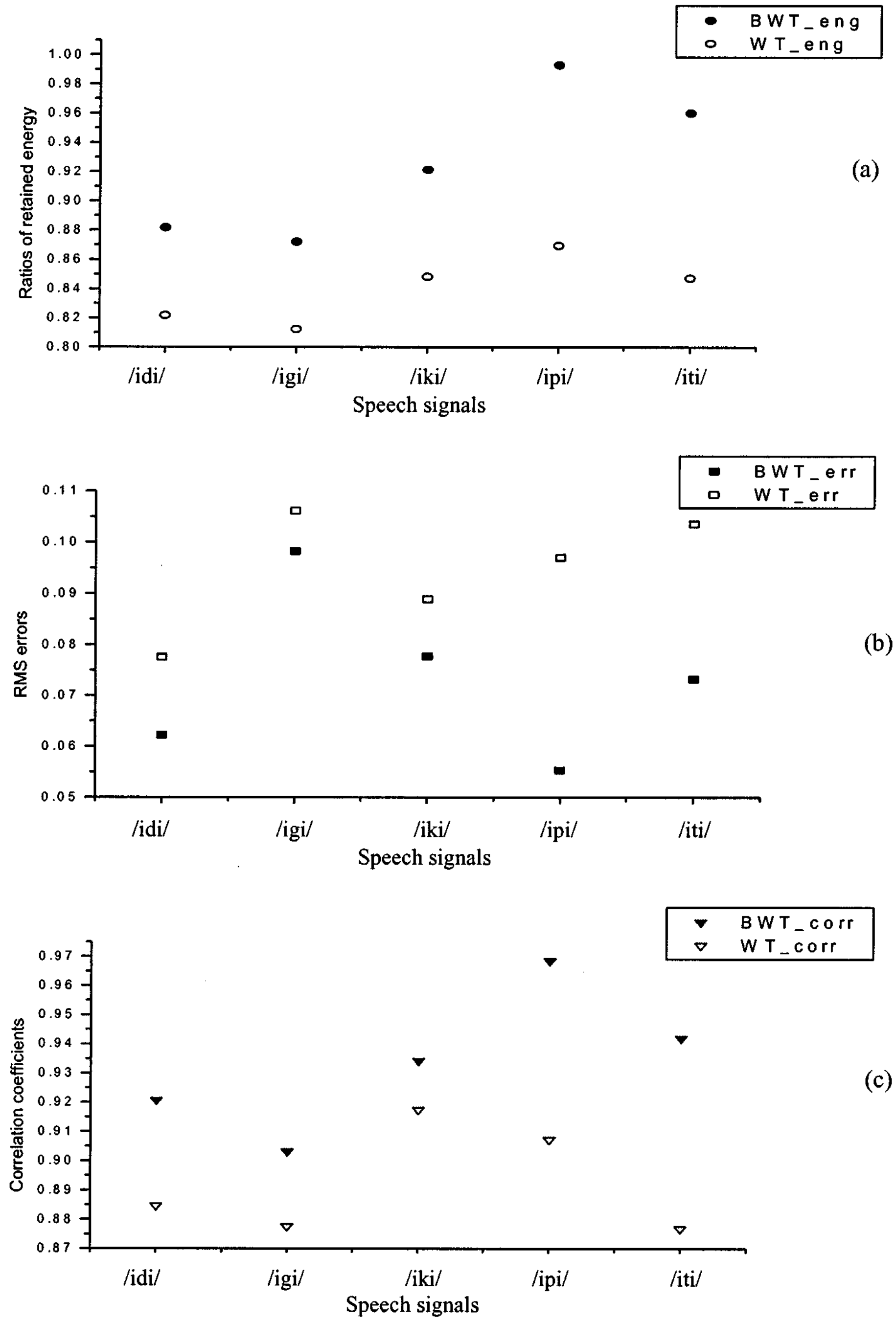

Fig. 5. The results of BWT application to speech signals, /idi/, /igi/, /iki/, /ipi/, and /iti/, in cochlear implants. (a) Ratios of retained energy. (b) The rms errors. (c) Correlation coefficients between reconstructed signals and original signals. Open and solid circles, squares, and triangles represent the ratios of retained energy, rms errors and correlation coefficients for BWT and WT, respectively.

1) A new parameter T, controlled by the signal instantaneous amplitude and its first-order differential, is introduced into the WT. With this additional parameter, the adaptive adjustment of resolution has been implemented in two dimensions.
2) BWT is a nonlinear transform that has high sensitivity and frequency selectivity.
3) BWT represents speech signals with more concentrated energy distribution than those of WT.
4) The inverse transform of BWT exists, which makes possible the utilization of BWT in signal coding and decoding.

References

[1] D. L. Jones and T. W. Parks, "A high resolution data-adaptive time–frequency representation," *IEEE Trans. Acoust. Speech, Signal Processing*, vol. 38, pp. 2127– 2135, Dec 1990.

[2] M. L. Hilton, "Wavelet and wavelet packet compression of electrocardiograms," *IEEE Trans. Biomed. Eng*, vol. 44, pp. 394–402, May 1997.

[3] U. Wiklund, M. Akay, and U. Niklasson, "Short-term analysis of heart-rate variability of adapted wavelet transforms," *IEEE Eng. Med. Biol. Mag.*, vol. 16, no. 5, pp. 113–118, 1997.

[4] D. Pollen and S. Ronner, "Phase relationships between adjacent simple cells in the visual cortex," *Science*, vol. 212, pp. 1409–1411, 1981.

[5] B. C. Moore, "Psychophysical tuning curves measured in simultaneous and forward masking," *J. Acoust. Soc. Amer.*, vol. 63, no. 2, pp. 524–532, 1978.

[6] T. Gold, “Hearing II: The physical basis of the action of the cochlea,” in *Proc. Roy. Soc. Lond.*, vol. 135, 1948, pp. 492–498.
[7] D. T. Kemp, “Stimulated acoustic emissions from within the human auditory system,” *J. Acoust. Soc. Amer.*, vol. 64, no. 5, pp. 1386–1391, 1978.
[8] R. Probst, “A review of otoacoustic emissions,” *J. Acoust. Soc. Amer.*, vol. 89, no. 5, pp. 2027–2066, 1991.
[9] D. J. Lim and F. Kalinec, “Cell and molecular basis of hearing,” *Kidney Int. Suppl.*, vol. 53, pp. s104–s113, 1998.
[10] L. Zheng, Y. T. Zhang, F. S. Yang, and D. T. Ye, “Synthesis and decomposition of transient-evoked otoacoustic emissions based on an active auditory model,” *IEEE Trans. Biomed. Eng*, vol. 46, pp. 1098–1106, Sept. 1999.
[11] J. Yao, Y. T. Zhang, and L. Zheng, “Modeling of otoacoustic emissions,” in *Proc. 3rd Int. Workshop Biosignal Interpretation*, Chicago, IL, 1999, pp. 198–201.
[12] J. Yao and Y. T. Zhang, “Cochlear is an inhomogeneous, active and nonlinear model,” in *Proc. IEEE EMBS/BMES*, Atlanta, GA, 1999, p. 1031.
[13] D. D. Greenwood, “A cochlear frequency-position function for several species—29 years later,” *J. Acoust. Soc. Amer.*, vol. 87, no. 6, pp. 2592–2605, 1990.
[14] C. Giguere and P. C. Woodland, “A computational model of the auditory periphery for speech and hearing research: II Descending paths,” *J. Acoust. Soc. Amer.*, vol. 95, no. 1, pp. 343–349, 1994.
[15] A. E. Hubbard and D. C. Mountain, “Models of the cochlea,” in *Auditory Computation*. New York: Springer, 1995.
[16] J. Yao and Y. T. Zhang, “The admissible condition of bionic wavelet transform and its inverse transform,” in *Proc. IEEE-EMBS Asia-Pacific Conf. Biological Engineering*, Hang Zhou, China, 2000, pp. 804–805.
[17] C. P. Behrenbruch, “SNR improvement, filtering and spectral equalization in cochlear implants using wavelet techniques,” in *2nd Int. Conf. Bioelectromagnetism*, Melbourne, Australia, 1998, pp. 61–62.
[18] K. B. Nie, N. Lan, and S. K. Gao, “Implementation of CIS speech processing strategy for cochlear implants by using wavelet transform,” in *Proc. ICSP*, 1998, pp. 1395–1398.
[19] J. Yao and Y. T. Zhang, The application of bionic wavelet transform to speech signal processing in cochlear implants, submitted for publication.

Jun Yao (S’98) was born in China, 1972. She majored in biomedical engineering and received the B.E. degree in 1995 from the Chongqing University, China, and M.E. degree in 1998 from the Southeast University, China. She is currently a Ph.D. degree candidate in the Department of Electronic Engineering, the Chinese University of Hong Kong, Hong Kong. Her Ph.D. research deals with the modeling of the ear based on otoacoustic emissions and its application to time–frequency signal processing.

Her current research interests include system modeling, signal processing especially on time–frequency analysis, artificial neural networks, and cochlear implants.

Yuan-Ting Zhang (M’90–SM’93) received the M.Sc. degree from Shan-Dong University, China, in 1981 and the Ph.D. degree from the University of New Brunswick, Fredricton, Canada, in 1990.

He was a Postdoctoral Fellow and Adjunct Assistant Professor in the University of Calgary, Calgary, Canada from 1989 to 1994. He joined the Chinese University of Hong Kong (CUHK), Hong Kong, as a Lecturer in 1994; became an Associate Professor in 1996; and serves currently as the Director of Joint Research Centre for Biomedical Engineering established between Zhejiang University and CUHK in 1999. He has authored and co-authored over 200 technical publications in areas as wide ranging as electromechanophysiological modeling, biosignal processing, medical devices, biomedical sensors, neuroengineering, and biomedical information engineering including tele-health and E-medicine.

He served as the Technical Program Chair of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology (EMBS), and a member of Editorial Boards of several journals and books including the Book Series of Biomedical Engineering published by the IEEE Press. He is currently the Vice-President of the IEEE Engineering in Medicine and Biology Society, and the Chairman of Biomedical Division of Hong Kong Institution of Engineers.