TELKOMNIKA, Vol. 11, No. 12, December 2013, pp. 7407~7414
e-ISSN: 2087-278X

An Auditory Model of Improved Adaptive ZCPA

Jinping Zhang[1], Yungong Li*[2], Hongqi Ni[1], Yanchun Zhao[1], Yaoqiang Zai[1]
[1]School of Mechanical Engineering, Shenyang University of Chemical Technology, Shenyang 110142, Liaoning, China
[2]School of Mechanical Engineering & Automation, Northeastern University, Shenyang 110189, Liaoning, China
*Corresponding author, e-mail: 66278038@qq.com

Abstract

An improved ZCAP auditory model with adaptability is proposed in this paper, and the adaptive method designed for ZCPA model is suitable for other auditory model with inner-hair-cell sub-model. The first step in the implement process of the proposed ZCPA model is to carry out the calculation of inner product between signal and complex Gammatone filters to obtain important frequency components of signal. And then, according to the result of the first step, the parameters of the basilar membrane sub-model and frequency box are automatically adjusted. Lastly an auditory model is built, and the final output is auditory spectrum. The results of numerical simulation and experiments have showed that the proposed model could realize accurate frequency selection, and the auditory spectrum is more distinctly than that of conventional ZCPA model. Moreover, the proposed model can completely avoided the influence of the number of filter so that the shape of auditory spectrum is steady, and the data quantity is small. Therefore the proposed model can be used as feature data in intelligent recognition and has greater practical applicability.

Keywords: *auditory model, adaptive signal analysis, frequency decomposition, features*

1. Introduction

Auditory model [1-3] is a signal analysis system which takes simulating the physiological structure and working mechanism of human auditory system as main characteristic, it have showed good working performance [4, 5] in those domains such as speech recognition, noise evaluation and features extraction of mechanical vibration. Auditory model mainly includes two components: auditory periphery and auditory centre model, among them, the core of auditory periphery model is cochlea basilar membrane model, its essence is a group of bandpass filter which is used to do bandpass filter to signal, this is the first work of whole auditory model and has larger effect on final output of model. Meanwhile, the parameters of basilar membrane such as filter number, each center frequency, band width, are changeless and uncorrelated with the features of analyzed signal, this will cause some shortages such that the analysis result depends on the parameter of basilar membrane model and has larger information redundancy.

The outer hair cell in human cochlea is an important component, basilar membrane can change self-dynamic properties on the self-regulation role of outer hair cell, thus which can select and filter to signal frequency automatically and enable the auditory system to have stronger nonlinear adaptive ability. Therefore, it is necessary to introduce the active regulative mechanism of outer hair inner into auditory model, which enables the auditory model to adaptively control the various parameters of basilar membrane filters, in order to realize the signal analysis of purposeful, optimal and higher flexible.

Based on the above reasons, the adaptive improvement ZCPA auditory model [6, 7] is proposed, it can built basilar membrane model according to signal self-characteristics in remaining all ZCPA advantages. The numerical simulation and experimental verification have showed that the model has larger improvementl in flexibility, accuracy and data quantity than those of the classical ZCPA model, and the proposed adaptive method can also be used to other auditory model.

Received July 23, 2013; Revised August 19, 2013; Accepted August 31, 2013

2. The Cochlea and ZCPA Model

2.1. The Basic Principle of Cochlea

The cochlea mainly consists of oval window, basilar membrane, Organ of Corti and auditory nerve, its function is to translate the mechanical energy of acoustic waves into neural coded signals. Among which, oval window receives the vibration of auditory ossicles and makes traveling wave vibration produce by the basilar membrane. Basilar membranes is a film which is dipping bath in lymph, and its width gradually increases from base to apex, as well as its elastic coefficient and damping also change, thus the acoustic waves of difficult frequency will produce the largest peak at the different position of basilar membrane, therefore, basilar membrane behaves the ability of bandpass filter. The vibration of basilar membrane will stimulate the inner hair cells of Organ of Corti for finishing the energy conversion from vibration stimulation to electrical stimulation, and introducing the electrical stimulation signal into the sensorineural auditory nerves. Between sensorineural auditory nerves and auditory center, the further treatment of electrical stimulation signal such as informational reduction, feature extracting and so on, will be completed by the neural network of lateral inhibition.

In addition, the outer hair cell is the very important component of cochlea, its input includes the response information of basilar membrane, the feedback information of afferent nerve (the nerve channel from cochlea to auditory center) and the control information of auditory center, the role of outer hair cells is to adjust the tension property of basilar membrane and change the electrochemical characteristics and inner environment of inner hair cell, that is the key factors of the nonlinear adaptive ability for human auditory.

2.2. ZCPA Auditory Model

ZCPA model is an auditory model which has concise calculation method, but doesn't take into account the role of outer hair cell, its implementation process is shown as Figure 1, this model includes three main steps, namely basilar membrane bandpass filter, the feature extracting and feature information synthesis of inner hair cell and auditory nerve.

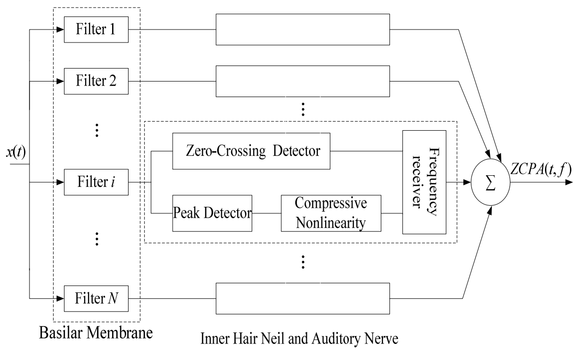


Figure 1. Principle Diagram of ZCPA Model

ZCPA model firstly simulates the function of basilar membrane that is to do band pass filter for signal $x(t)$ when the channel number is N , let the i th filter be $h(t,i)$, $h(t,i)$ center frequency is determined by i , and the filter output is $y_1(t,i)$ .then we detect all upward-going zero-crossing points, the time interval and the wave peak between zero-crossing points in each channel filter signal called $y_1(t,i)$, let the time interval of i th signal between l th and $l+1$th zero-crossings of $y_1(t,i)$ be ΔT_{il} , and peak be p_{il} . Let the normalization processing result of p_{il} still be $\overline{p}_{il}$, then $\overline{p}_{il}$ will be compressed nonlinearly [7], that is:

$$s(\overline{p}_{il}) = \frac{2}{1+\exp(-\gamma \cdot \overline{p}_{il})} - 1 \quad (1)$$

Before doing statistical analysis about ΔT_{il} and $s(\overline{p}_{il})$, M frequency intervals named frequency box [7] need to be divided. The corresponding output of frequency receiver is described as:

$$y_2(t,m,i) = \sum_{l=1}^{Z_i-1} \delta_{mil} s(p_{il}) \quad (m = 1,2,\cdots,M) \tag{2}$$

Where Z_i is the total number of zero-crossing, m is the order number of frequency box, each m is corresponding to a frequency range, δ_{mil} is Kronecker operator, if $f_{il} = 1/\Delta T_{il}$ falls into the m th frequency box, then $\delta_{mil} = 1$, else $\delta_{mil} = 0$. The final output of ZCPA model is:

$$zcpa(t, f_m) = \sum_{l=1}^{N} y_2(t,m,i) \quad (m = 1,2,\cdots,M) \tag{3}$$

Where f_m is the center frequency of m th frequency box, $zcpa(t, f_m)$ is usually called as auditory spectral.

2.3. Gammatone Filter Characteristics

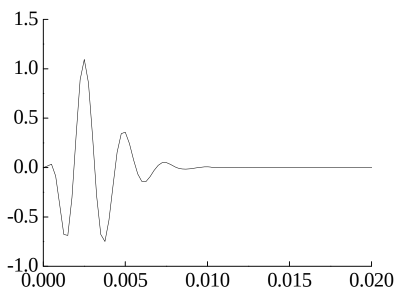
Now, most auditory models use the Gammatone filter bank (hereinafter called as GT filter) to simulate the bandpass filter function of cochlea basilar membrane, The time-domain expression [8] of gammatone filter is:

$$h(t,i) = B^n t^{n-1} e^{-2\pi Bt} \cos(2\pi f_i t + \phi_i) \tag{4}$$

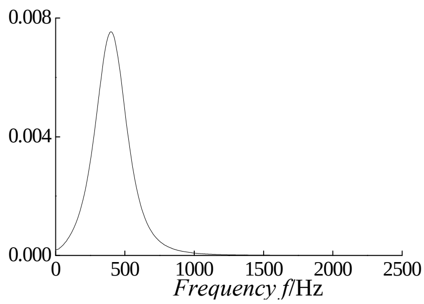
Where, f_i is center frequency, n is filter rank, it can well simulate the characteristics of basilar membrane as n=4. Phase Φ_i is generally taken as zero, the calculation formula of B is described as:

$$B = 1.019 \cdot \mathrm{ERB}(f_i) \tag{5}$$

Here, $\mathrm{ERB}(f_i)$ is the equivalent rectangular bandwidth of filter and $\mathrm{ERB}(f_i) = 1.019 \cdot (24.7 + 0.108 f_i)$. When the center frequency is 400Hz, the time domain wave and amplitude spectrum of the filter are shown in Figure 2.



(a) Time domain wave



(b) amplitude spectrum

Figure 2. The Filter Wave and Amplitude Spectrum when Center Frequency is 400Hz

From the Figure 2 it can be found that GT filter got by Equation (4) and (5) behaves better locality of time domain, but poorer locality of frequency domain. Meanwhile, according to the real characteristics of basilar membrane, some auditory model usually sets the center frequency of GT filter taking this way as logarithm uniform distribution along frequency axis. Figure 3 is the distribution situation of each filter amplitude spectrum when sampling frequency is 1000Hz as well as N=18, in which the abscissa and ordinate are both logarithmic coordinates. Because the setting mode and the self-characteristics of GT filter bank, the frequency overlap is larger between each pair of filters, thus will increase the information redundancy and frequency ambiguity of filter result.

Considering that the frequency resolution of human ear can reach 1Hz and choose more accurate frequency autonomously, therefore, it is necessary to introduce adaptive function and active control mechanism in auditory model, that will improve the information processing function of model sequentially.

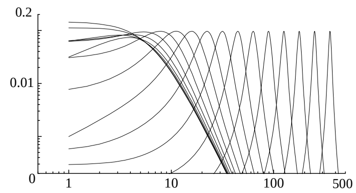


Figure 3. The Amplitude Spectrum of Gammatone Filter Bank when N=18

3. Adaptive Improvement of ZCPA Model

The purpose of adaptive improvement for ZCPA model is to make each center frequency and bandwidth of GT filters bank to coordinate autonomously with the structure of signal frequency, and can realize accurate location of frequency and frequency selection, which has two core problem, one is to rapidly pre-scan the signal for conforming clearly the frequency structure of signal, another is to formulate the principles of the parameters control for GT filter bank and frequency box. The implement principle of improvement ZCPA model is shown in Figure 4.

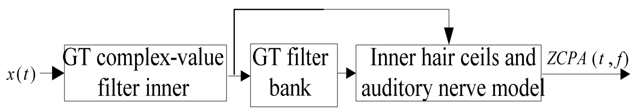


Figure 4. The Principle of Adaptive ZCPA Model

3.1. The Methods of Signal Pre-scanning

The purpose of signal pre-scanning is to provide basis for adjusting the parameters of GT filter, the simplest method is to perform the Fourier transform of signal at each time segment, and the amplitude spectrum is taken as scanning result. But when the Fourier transform needs more signal points, which maybe lose the signal transient feature. Meanwhile, from the views of the feature waveform, the pre-scanning should as far as possible descript the similarity degree between the signal waveform of each frequency range and the corresponding GT filter waveform at certain period, this also can't be done by Fourier transform. So this paper presents the pre-scanning approach by using complex-valued GT filter to do inner product for signal.

Given complex-valued GT filter as:

$$h_c(t, f) = b^n t^{n-1} e^{-2\pi bt} e^{j2\pi ft} \tag{6}$$

Where, the calculation method of b is:

$$b = \frac{1}{\eta}(24.7 + 0.108f) \tag{7}$$

Here η is parameter, the filter bandwidth can be control by adjusting the value of η.

Because the calculation of non-stationary signal is usually segmentation namely each paragraph is assumed as stationary signal, therefore, for expressing simply, the signal $x(t)$ is assumed as stationary, and its length is denoted as T. The inner product between $h(t,f)$ and $x(t)$ is described as:

$$X(f) = \frac{1}{T}\int_0^T x(t)h_c(t,f)dt \tag{8}$$

And then the modulus of $X(f)$ named $D(f)$ is shown as:

$$D(f) = |X(f)| \tag{9}$$

Finds all the extreme points of $D(f)$ which amplitude are bigger than the threshold θ_a, and let the extreme points number be I, the amplitude and its corresponding frequency of i th extreme points be A^i and f^i, respectively. Meanwhile, the frequency interval between each adjacent extreme points is calculated, named the frequency interval between i th and $i+1$ th as Δf^i, meanwhile, $\Delta f^0 = f^1$ and $\Delta f^i = f_s/2 - f^i$.

3.2. Parameters Control of GT Filters Bank and Frequency Box

The parameters of I and f^i as well as Δf^i are gotten after pre-scanning, it is obvious that the adaptive control of GT filters bank parameters depend on this three kinds of data. In which, I is the number of needed filter, f^i is the center frequency of each filter, Δf^i determines the bandwidth of each filter, the bandwidth of the filter center frequency being f^i can be described as:

$$b^i = \frac{1}{2}\min[\Delta f^{i-1}, \Delta f^i] \tag{10}$$

The filters bandwidth gotten from equation (10) only depends on the frequency interval between adjacent frequency components, thus that will not cause the omissions of signal, and also doesn't cause effectively analytic signal because too big bandwidth (especially the high frequency part).

When sets the frequency box, firstly segments several frequency boxes along frequency axis in the same manner as depicted in Section 2.2, then embeds I numbers frequency boxes related to f^i, the corresponding frequency range is $[f^i - \xi, f^i + \xi]$, here ξ is parameter. The above process can be shown as Figure 5.

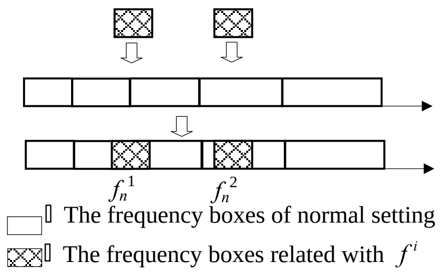


Figure 5. Frequency Boxes Setting Schematic Diagram

4. Model Validation
4.1. Simulation Validation
Let signal $x(t)$ be:

$$x(t) = \cos(2\pi \cdot 10 \cdot t) + \cos(2\pi \cdot 120 \cdot t) + 4\cos(2\pi \cdot 200 \cdot t) + 2\cos(2\pi \cdot 300 \cdot t) + \cos(2\pi \cdot 500 \cdot t)$$

For the sampling frequency f_s of 3000 Hz, model parameter η of 10, let the step size be $f_s/2000$ when calculates the inner product according to equation (8), and $\xi = 2 \cdot f_s/2000$, $\theta_a = 0.02\max[D(f)]$， $\gamma = 5$ when does nonlinear compression, let the initial setting number of frequency box be 20. The auditory spectrum getting by the presenting model in this paper to analyze $x(t)$ is shown in Figure 6. For to compare, we use the classical ZCPA model as described in Section 2.2 to analyze $x(t)$, the frequency number still take as 20, and the filter number N take for 16 and 40, respectively, then the gotten two auditory spectrum are shown in Figure 7. Contrasting Figure 6 and Figure 7, it can be found that this paper proposed model can be more accurate in describing the frequency structure of signal than that of the classical ZCPA model, and only 5 filters are used in the process of filter for basilar membrane. At the same time, the classical ZCPA model uses more filter number, its description to signal feature is inadequate precise, the frequency position is not accurate enough, and the frequency component with 300Hz couldn't be showed in Figure 7(a), it explains that the relations between the component number of signal frequency and filter number have greater influence on the analysis result.

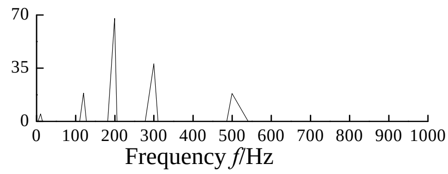


Figure 6. The Auditory Spectrum of the Proposed Model

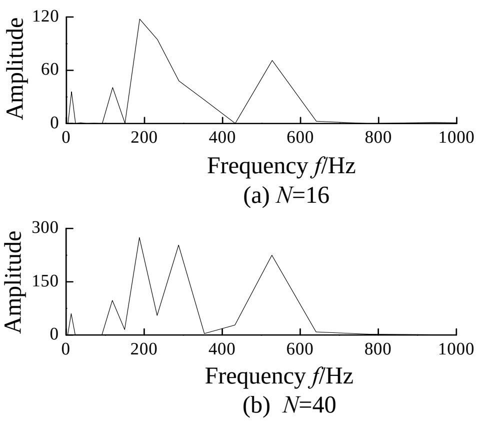


Figure 7. The Auditory Spectrum of Classical ZCPA Model

4.2. Experiment
Using the rotor test-bed to do rotor rubbing experiment, sampling frequency is 2000 Hz. The test-bed is shown in Figure 8, the detected wave and amplitude spectrum are shown as Figure 9.

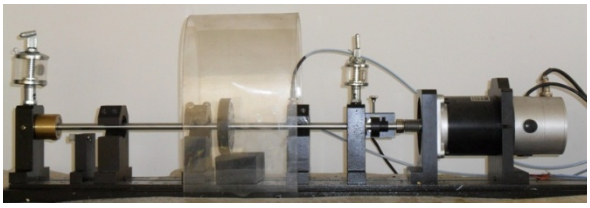
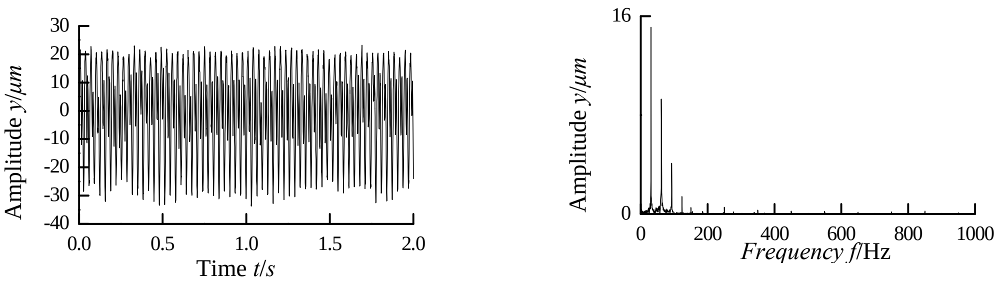


Figure 8. The Rotor test-bed

Using this paper proposed model and classical ZCPA model to do analysis for the signal showed in Figure 9, respectively. Parameter setting uses the same manner as described in Section 4.1, in which, the filter number of basilar membrane for classical ZCPA model is 40, and the results are shown in Figure 10 and Figure 11, respectively. Comparing Figure 9(b), Figure 10, and Figure 11, it can be found that the presented model can realize finer frequency decomposition for the engineering signals with certain randomness, but classical ZCPA model can only characterize out the fundamental frequency and second harmonic component, that indicates this paper proposed model has stronger practicality, the results data of final calculation are far less than those of the amplitude spectrum as showed in Figure 9(b) (the data of auditory spectrum of Figure 10 is 13). At the same time, it is need to explain that the amplitude of auditory spectrum has bigger difference between Figure 10 and Figure 11, this is caused by the difference of the used filter bandwidth and number of the two model.



(a) Time domain wave (b) amplitude spectrum

Figure 9. The Waveform and Amplitude Spectrum of Rotor Rubbing Vibration Signal

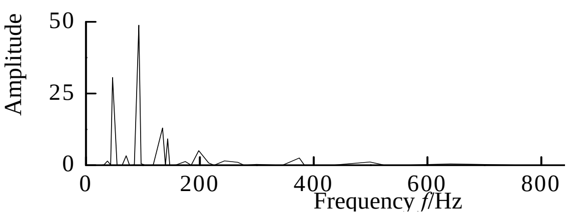


Figure 10. Auditory Spectrum of the Proposed Model

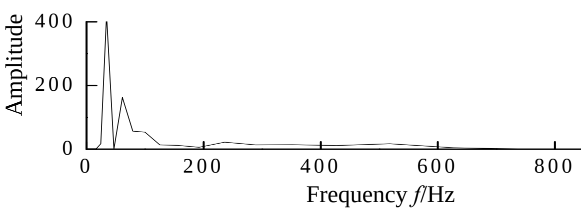


Figure 11. Auditory Spectrum of the Classical ZCPA

5.Conclusion

According to the functional characteristic of outer hair inner, this paper puts forward an adaptive ZCPA auditory model, comparing with classical ZCPA model, the proposed model in this paper has the following features. Firstly, the proposed model can adaptively determine the filter member and each filter center frequency as well as bandwidth of membrane model, thus can avoid the blindness of membrane building model in classical model. Seconly, the frequency segments relating to the analyzed signal are adaptively imbedded in frequency boxes to improve the focusing ability of the main frequency components in auditory spectrum. Thirdly, in the relatively smaller data amount, the proposed model has more accurate and clear description to signal, and don't exist the information redundancy. The proposed model in this paper has higher flexible and initiative as well as the application prospect of certain actual, and suitability in the related situation such as machinery fault feature extraction and so on, also can be the principal character of intelligent identification.

Acknowledgement

This work was supported by Science and Technology Research Project of Department of Education of Liao Ning Province (L2010442), National Natural Science Foundation of China (50805021), Fundamental Research Funds for the Central Universities (N100403004).

References

[1] Serajul H, Roberto T, Anthony Z. A temporal auditory model with adaptation for automatic speech recognition. *ICASSP*. 2007; 1141-1144.
[2] Klapuri A. Multipitch analysis of polyphonic music and speech signals using an auditory mode. *IEEE Transac- tions on speech and audio processing*. 2008; 16(2): 255- 266.
[3] Chu W, Champagne B. A Simplified Early Auditory Model with Application in Speech/Music Classification. *IEEE CCECE/CCGEI*. Ottawa. 2006; 775-778.
[4] LI Yungong, ZHANG Jinping, DAI Li, ZHANG Zhanyi, LIU Jie. Auditory-Model-Based Feature Extraction Met- hod for Mechanical Faults Diagnosis. *Chinese Journal of Mechanical Engineering*. 2010; 21(3): 391-397.
[5] LI Yungong, ZHANG Jinping, DAI Li, ZHANG Zhanyi, LIU Jie. Auditory spectrum of mechanical vibration signal and its characteristics. J*ournal of Vibration and Shock*. 2010; 29(11): 204-208.
[6] Serajul H, Roberto T, Anthony Z. Perceptual features for automatic speech recognition in noisy environments. *Speech Communication*. 2009; 51(1): 58-75.
[7] LI Yungong, ZHANG Jinping, DAI Li, ZHANG Zhanyi, LIU Jie. Study on the ZCPA-auditory-model-based Method of Feature Extraction for Mechanical Faults Diagnosis. *China Mechanical Engineering*. 2009; 20(24): 2988-2992.
[8] Schluter R, Bezrukov L, Wagner H, Ney H. Gammatone features and feature combination for large vocabu- lary speech recognition. *ICASSP*. 2007; 679- 652.