A Feature Extraction Method based on Combined Wavelets Filter in Speech Recognition

Xueying Zhang, Ying Sun
College of Information Engineer
Taiyuan University of Technology
Taiyuan, China
zhangxy@tyut.edu.cn; tyutsy@163.com;

Wenjun Hou
Automation School
Beijing University of Posts and Telecommunications
Beijing, China
wenjunh2113@263.net;

***Abstract*—This paper used wavelet theory in noise-robust feature extraction of speech recognition and introduced two kinds of feature extraction methods based on Gauss wavelet filter and combined wavelets filter. The Gauss wavelet filter and combined wavelets filter with critical frequency bands are obtained by studying human auditory characteristic. Wavelet has flexible characteristic in choosing frequency, the key is making certain the scale parameter. This paper studied the choosing method of scale parameter in designing the two kinds of wavelet filter. The methods used new feature and original feature were simulated. The RBF neural net is used in training and recognition course. The results showed that new feature had higher recognition rate and better robustness than traditional feature.**

***Keywords*—robustness, feature extraction, speech recognition, wavelet transform, filter**

I. Introduction

The purpose of speech recognition is to make the computer understand human language. For speech recognition, it is not a dream that people can communicate with machine directly [1]. Usually the speech recognition system includes front-end process, feature extraction and recognition network three parts. Among them, feature extraction is a key, because better feature is good for improving recognition rate. Recently, as the development of research, the speech feature extraction based on auditory model catches more sight. This is because auditory model is close to the processing of human ears. So the feature based on auditory model can represent speech essentially.

Wavelet transform is a new method of signal processing in recent years. It has better location characteristic and the resolution in time-frequency domain is changed. These characteristics of wavelet make it become an effective way for analyzing non-stationary signal. Speech signal is just one kind of non-stationary signal. This paper presents the improved zero-crossings with peak-amplitudes (ZCPA) feature [2] which uses the Gauss wavelet filters and combined wavelets filter instead of the FIR filters in ZCPA preprocessing. The experiment showed that new feature had better robustness than ZCPA feature.

II. The Feature Extraction of ZCPA

The ZCPA feature extraction system consists of a bank of band-pass filters, zero-crossing detector, peak detector, nonlinear compression, frequency receiver and normalization of the time and amplitude [3].

The feature extraction of ZCPA uses zero-crossing interval to represent signal frequency information and amplitude value to represent intensity information, finally frequency information and amplitude information is combined to form the complete feature output. Figure 1 shows the block diagram of ZCPA model.

III. A Feature Extraction Method Based on Gauss Wavelet Filter

According to human auditory characteristic, human ears have auditory masking effect. The frequency ranged from 20Hz to 16000Hz could be separated into 24 critical bandwidths [4]. We choose suitable parameter to build new filters in order to make the new feature be more closely with the essence of speech and make the new bandwidths be fit to critical bandwidths. This paper first introduces a feature extraction method based on Gauss wavelet filter in speech recognition. New feature is obtained by instead Gauss wavelet filters of original FIR filters in ZCPA. This new feature is called GZCPA feature.

Define Gauss wavelet as this equation [5]: $\psi(t)=\frac{1}{2\sqrt{\pi\alpha}}e^{-\frac{t^2}{4\alpha}}, \quad \alpha>0$

Set $\alpha=\frac{1}{2}$, we can get $\psi_1(t)=\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}$. By normalizing it, we get $\psi_2(t)=e^{-\frac{t^2}{2}}$, then do scale transformation for $\psi_2(t)$, the result is:

$$\psi(t)=\frac{1}{2\sqrt{\pi\alpha}}e^{-\frac{t^2}{4\alpha}}, \quad \alpha>0$$

Set $\alpha=\frac{1}{2}$, we can get $\psi_1(t)=\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}$. By normalizing it, we get $\psi_2(t)=e^{-\frac{t^2}{2}}$, then do scale transformation for $\psi_2(t)$, the result is:

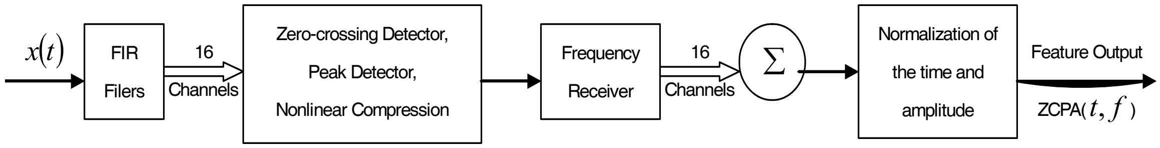

Figure 1. Block Diagram of the Zero-crossing with Peak Amplitudes (ZCPA) Model

$$\psi_s(t)=\frac{1}{s}\psi_2\left(\frac{t}{s}\right)=\frac{1}{s}e^{-\frac{t^2}{2s^2}} \tag{1}$$

Where s is the scale factor, changing s can change the width of wavelet frequency window. In the case that the shape of the window is not change, if the central frequency of the wavelet frequency window is 0Hz, and the interested frequency composition is f_0 , the center of the frequency window can be shifted to f_0 by multiplying $\psi_s(t)$ with shift factor $e^{j2\pi f_0 t}$. Based on above principle, make sure the center frequencies of the windows by the center frequencies of 16 critical bandwidths. Then choosing the proper scale factor s , the design of Gauss wavelet filters can be completed. Following conditions should be satisfied for selecting scale factor s : (1) Frequency windows should have some overlapping between two adjacent windows. After some experiments, setting the value of cross-over point is 0.8. (2)Width of wavelet frequency windows should be the same with corresponding bandwidth.

The final s values are listed in Table 1.

Making Fourier transform for $\psi_s(t)e^{j2\pi f_0 t}$, we could get the frequency shifted Gauss wavelet equation in frequency domain:

$$\begin{aligned}\Psi(f) &= \int_{-\infty}^{+\infty}\psi_s(t)\cdot e^{j2\pi f_0 t}\cdot e^{-j2\pi ft}\,dt \\ &= \int_{-\infty}^{+\infty}\frac{1}{s}e^{-\frac{t^2}{2s^2}}\cdot e^{j2\pi f_0 t}\cdot e^{-j2\pi ft}\,dt \\ &= \sqrt{2\pi}\,e^{-2\pi^2 s^2 (f-f_0)^2}\end{aligned} \tag{2}$$

Because computers can only operate discrete signal, it is necessary to make $\Psi(f)$ discrete. Set $f_s = 1/T_s$, f_s is sample frequency, N is the length after adding zero, here the length is 256 samples. Then the equation (3) is the form of $\Psi(f)$ after discreteness:

$$\Psi(N-i)=\Psi(i-1)=\Psi\left(i\cdot\frac{f_s}{N}\right) \qquad i=1,2,\cdots\frac{N}{2} \tag{3}$$

Making wavelet transform for input speech signal as (4), where $X(l)$ is the spectrum of $x(t)$:

$$W(l)=\Psi(l)\cdot X(l) \tag{4}$$

The essence of equation (4) is band-pass filter, it extracts the information that is determined by the band pass characteristic of wavelet $\psi_s(t)$, and also the center frequency of wavelet $\psi_s(t)$ is f_0 .

Figure 2 is the frequency response of 16 Gauss wavelet filters.

IV. A FEATURE EXTRACTION METHOD BASED ON COMBINED WAVELETS FILTER

A. Theory of Combined Wavelets Filter

From Figure 2 we can see the frequency response of filters is comparative sharp. This is good for narrow-band signal. But the desired frequency respond of filters should have smoother top. So this paper introduced the theory of combined wavelets filter.

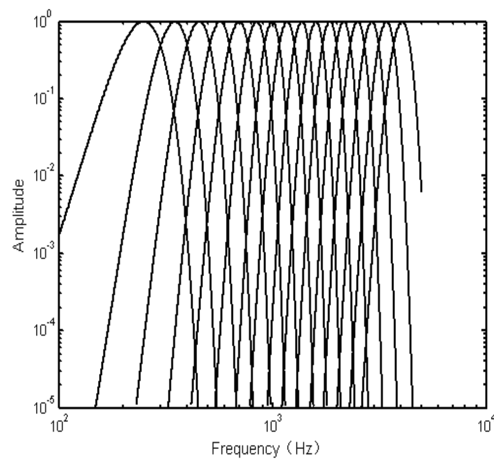

Figure 2. The Frequency Response of GaussWavelet Filters

If $\psi(t)$ is the basic wavelet which satisfies the 'admissible condition', similar to the Gauss wavelet filter, the interested frequency component can be got by multiplying $\psi(t)$ with shift factor $e^{j2\pi f_0 t}$.

Combined wavelets are got by multiplying frequency shifting and superposing single wavelet. If the frequency center of $\psi(t)e^{j2\pi f_k t}$ is f_k , after superposing single wavelet whose frequency is shifted, the combined wavelets are:

$$\begin{aligned}\psi_g(t) &= \sum_{k=1}^{n}\psi(t)e^{j2\pi f_k t} \\ &= \psi(t)\left(e^{j2\pi f_L t}+\cdots\cdots+e^{j2\pi(f_L+n\Delta f)t}+\cdots\cdots+e^{j2\pi f_H t}\right)\end{aligned} \tag{5}$$

$$k=0,1,2\cdots,n; \quad f_H=f_L+n\Delta f$$

TABLE I. 16 CENTRAL FREQUENCIES OF CRITICAL BANDWIDTH f_i, THE BANDWIDTH Δf AND S VALUE

Bark No.	3	4	5	6	7	8	9	10
f_i (Hz)	250	350	455	570	700	845	1000	1175
Δf (Hz)	100	100	110	120	140	150	160	190
s value	0.0021	0.0021	0.0019	0.00173	0.0015	0.00141	0.00132	0.00113
Bark No.	11	12	13	14	15	16	17	18
f_i (Hz)	1375	1600	1860	2160	2510	2925	3425	4050
Δf (Hz)	210	240	280	320	380	450	550	700
s value	0.001	0.00089	0.00076	0.000665	0.00056	0.00047	0.000388	0.000303

Where f_L and f_H are the center frequency of the first and final wavelet, Δf is the frequency center interval, respectively.

In this way, according to critical frequencies in Table 1, combined wavelets filter can be got. And each filter is combined with $n+1$ single wavelet.

B. A Feature Extraction Method Based on Combined Wavelets Filter

Similar to Gauss wavelet filter, we build a group of combined wavelets filters by the combined wavelets theory to improve the front-end filters in ZCPA. The bandwidths of new filters are also followed critical bandwidths. Obtain new feature by instead combined wavelets filters of original FIR filters in ZCPA. This new feature is called CWZCPA feature. The calculation steps of improved algorithm are as follows:

(1) Frame the input speech signal, the length of each frame is 20ms; the frame-shift is 10ms.

(2) A FFT is computed for framed speech signal to obtain the frequency information of speech signal in every frame.

(3) Choose proper basic wavelet for combined wavelets, in this paper the basic wavelet is Gauss wavelet.

(4) Choose proper parameter according to combined wavelets theory. Take the first filter of 16 critical bandwidths filter for example. The center frequency of this band is 250Hz, the bandwidth is 100Hz, so the frequency range is 200Hz to 300Hz. According to equation(5), set $f_L = 200$ Hz, $f_H = 300$ Hz. For making the frequency respond more smooth, choose $\Delta f = 50$ Hz, and transform equation (5): $n = \frac{f_H - f_L}{\Delta f}$, get $n = 2$. This means the first critical frequency band is combined with $n+1 = 2+1 = 3$ wavelets, the center frequency of the first single wavelet is 200Hz, the second one is 250Hz, the last one is 300Hz. Combined these three single wavelets, the frequency respond of the first critical frequency can be obtained. The way to combine last 15 filters similar as the first one. The only difference is the center frequency of the first and final wavelet in each band. Figure 3 is the frequency response of 16 combined wavelets filters.
(5) Obtain the final feature by putting the output of 16 filters into the feature extraction system of ZCPA.

(6) Send the feature parameter into RBF network for training and testing; at last the recognition results can be obtained.

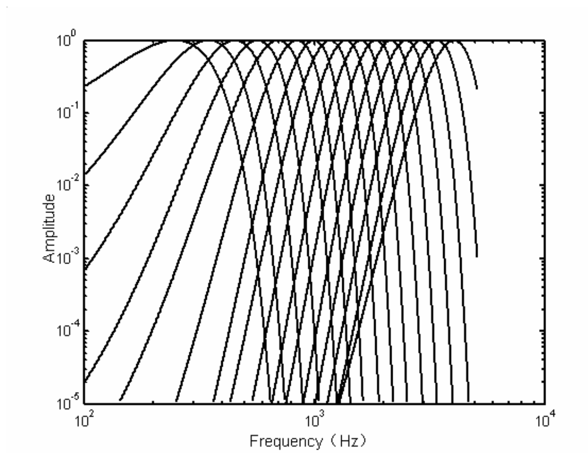

Figure 3. The Frequency Response of Combined Wavelet Filters

V. THE EXPERIMENT RESULTS AND ANALYSIS

In the experiment, the recognition network is RBF. The recognition object is speaker-independent, isolated-words. The vocabulary table is 10 words, 20 words, 30words, 40 words and 50 words. 16 speakers uttered the words 3 times, and use 9 people's utterance as training set, 7 people's utterance as testing set. The additive noise is white Gaussian noise; the SNR is divided into five degree conditions ranging from clean to 15dB. Each recognition result of ZCPA, GZCPA and CWZCPA are listed in Table 2.

The experiment results show that after using Gauss wavelet filters and combined wavelets filters, most of the recognition rate has increased. Comparing the two methods, the frequency respond of CWZCPA is much smoother than GZCPA and the recognition rate of CWZCPA is higher than the one of GZCPA. From the results we can also find the recognition rate of improved methods are increased more obviously in smaller vocabulary, but when the vocabulary becomes larger, the degree of increment decreased. This is because that the training set is not large enough or Gauss wavelet is not suitable for large vocabulary. So the main goal of next work is studying new wavelet to improve the recognition result.

TABLE II. RECOGNITION RATE OF ZCPA, GZCPA AND CWZCPA(%)

Word No.	Feature Type	SNR（dB）				
		15	20	25	30	clean
10	CWZCPA	89.5	92.4	95.2	96.2	95.7
	GZCPA	90.0	92.4	93.8	94.8	95.2
	ZCPA	82.4	84.8	85.2	87.1	87.1
20	CWZCPA	87.4	90.2	93.3	93.8	94.8
	GZCPA	86.0	89.0	92.1	91.9	93.8
	ZCPA	82.6	84.8	86.7	87.1	87.6
30	CWZCPA	88.1	90.5	92.7	93.5	94.8
	GZCPA	87.0	90.6	91.9	92.7	94.0
	ZCPA	83.1	85.1	88.4	89.2	89.0
40	CWZCPA	87.1	89.6	92.6	93.3	93.9
	GZCPA	85.6	89.3	91.4	92.6	93.7
	ZCPA	84.0	85.9	89.3	90.4	91.6
50	CWZCPA	85.7	88.3	91.3	92.8	93.5
	GZCPA	84.7	88.0	90.0	91.6	92.5
	ZCPA	85.0	88.1	90.5	91.1	92.3

ACKNOWLEDGMENT

The project is sponsored by the Natural Science Foundation of China (No.60472094), Shanxi Province Scientific Research Foundation for Returned Overseas Chinese Scholars (No.2006-28), Shanxi Province scientific Research Foundation for University Young Scholars ([2004] No.13), Shanxi Province Natural Science Foundation. (No.20051039).The authors gratefully acknowledge them.

REFERENCE

[1] Gang Zhang, Xueying Zhang, Jianfen Ma, *Speech Processing and Coding*, Publishing House of Ordnance Industry, Beijing, China, 2000, pp.98-99.

[2] Doh-suk Kim, Soo-Young Lee, Rhee M.Kil, “Auditory Processing of Speech Signal for Robust Speech Recognition in Real-World Noisy Environments”, *IEEE Transactionson speech and audio processing,* vol.7, no.1, pp.55-68, Jan. 1999. .

[3] Zhiping Jiao, Xueying Zhang, Shuyan Zhao, “A Robust Noise Feature Extraction Method Based on Auditory Model in Speech Recognition”, *Journal of Taiyuan University of Technology,* China,vol.36, pp.13-15, Jan. 2005..

[4] Zhiping Jiao, “Algorithm Research Based on The Improved ZCPA Speech Recognition Feature Extraction”, Master Degree Dissertation of Taiyuan University of Technology, China, pp.30-31, 2005..

[5] Jiancang Ma, Lei Luo, “Envelop Analysis Method Based on Wavelet Transform”, *Journal of Northwestern Polytechnical University*, China, vol.15, no.2, pp.249-253, May 1997..