

## Study of human identification by Electrocardiography frequency features

Wang Yarong<sup>1, a</sup>, Zheng Gang<sup>1, b</sup>

<sup>1</sup> School of computer and communication engineering, Tianjin University of Science and Technology, China

<sup>a</sup>494519703@qq.com, <sup>b</sup>kenneth\_zheng@vip.163.com

### Abstract

The paper proposed a human identification strategy based on ECG (electrocardiogram) frequency domain features. Fourier transform was used on ECG signal to get frequency domain features. The features contained frequency-domain signal slope, harmonic number, the magnitude gap, ratio of different frequency energy to total energy. Classifiers of correlation analysis and neural network were used in identification, and the accuracy of neural network reached 96.4%.

*Keywords: amplitude frequency characteristics; energy spectrum; the neural network*

### 1 Introduction

As important biomedical signals, ECG signal reflects the heart and the nature of life activities. ECG signal not only can be used in clinical diagnosis, but also can be applied to identify the characteristics of the identity. Simon explained the advantage of ECG which is used to the identification[1],for example,(1)The ECG signal is universal and easy to acquire;(2)The ECG signal has the long-term stability; (3) The ECG signal are different: because of the location、 size and anatomical structures of different people's heart, ECG signals are different from person to person; (4) ECG signal is not easy to be counterfeit: Although ECG signal has a long-term stability, it changes with the change of physical condition, age and body weight and so on. When the ECG signal changes, template can be updated cumulatively and becomes a kind of dynamic templates.

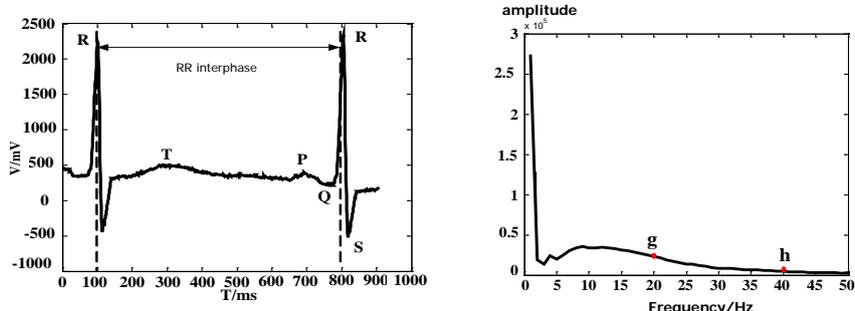


Figure 1 ECG signal waveform in time domain Figure 2 A single cycle of ECG signal amplitude spectrum

For a long time, the method of identity of the ECG signal was widely concerned. Biel proved the feasibility of identification using ECG signals for the first time [2]. Human identification by Electrocardiography time features such as amplitude of R wave、amplitude of T wave、QRS interphase、QT interphase and so on[3], as shown in figure 1, But the time domain waveform characteristics are more easily affected by the change of heart rate and depends on the automatic identification accuracy of ECG signal in time domain.

In recent years, there are many persons researching methods of the feature extraction which is not depend on reference point [4-9]; This Method is put forward in 2004 at the earliest. For example, Plataniotis used the method of AC/DCT, ECG signals after the pretreatment are cut into different windows, the data of every window is used to correlate analysis, DCT is used to Dimension reduction and won the recognition rate of 92.8% [10]. In 2008, Chan proposed a method based on wavelet transform for feature extraction, extracting wavelet coefficient as Characteristics of identity recognition and won the recognition rate of 100% for normal person and 81% for patients with arrhythmia [11].

Fourier transform was used on RR interphase of ECG signal to get frequency domain features. The features contained frequency-domain signal slope, harmonic number, the magnitude gap, ratio of different frequency energy to total energy. Respectively using the correlation coefficient and the neural network to identify characteristics and reached a better recognition rate.

## 2 Feature extraction on ECG signal in frequency domain

### The Fourier transform of ECG signal

The paper takes the RR interphase as a single cycle of ECG signals with  $X(t)$ , As shown in figure 1. Spectrum density function of  $X(j\omega)$  is the Fourier transform of  $X(t)$ , such as Eq.1. Amplitude spectrum represented by  $|X(j\omega)|$ , such as Eq. 2. Amplitude spectrum said the relationship between frequency and amplitude is shown in figure 2:

$$X(j\omega) = \int_{-\infty}^{\infty} f(t)e^{-j\omega t} dt = R(\omega) + jX(\omega) \quad (1)$$

$$|X(j\omega)| = \sqrt{R(\omega)^2 + X(\omega)^2} \quad (2)$$

### Character description

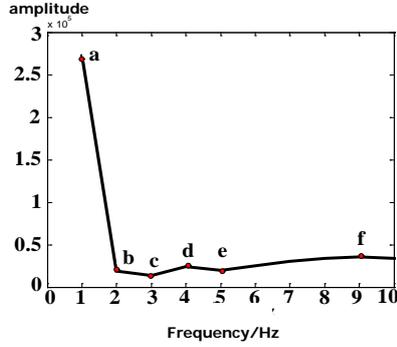


Figure 3 amplitude spectrum of ECG signal from 1Hz to 10Hz

(1) The a-b slope: “a” is the first point of the ECG signal.  $k_A$  is the slope of “a” to the next point, Such as Eq.3:

$$k_A = (y_{next} - y_a) / (x_{next} - x_a) \quad (3)$$

recording slope of the second point to the third point, the third point to the fourth point and so on. If the slope is not equal to  $k_A$ , then stop, such as “b”, is shown in figure 3. Assuming that the coordinate of point a and b are  $(x_a, y_a)$  and  $(x_b, y_b)$ , k is used to represent the slope of point a to b. Such as Eq.4:

$$k = (y_b - y_a) / (x_b - x_a) \quad (4)$$

(2) The number of harmonic: The number of peaks between point b and 20 Hz and then location of the peaks is found.

(3) The b-c slope: Recording the location of peaks and trough between point b and 20 Hz and “c” is the amplitude of the smallest number in the recording.

(4) The magnitude gap: Data in the recording is sorted in descending. Amplitude difference of the first and the second point, the second and the third point are got and so on, adding up the amplitude difference.

(5) The g-h slope: As figure 3 shows “g” is the amplitude of 20Hz and “h” is the amplitude of 40Hz and then the slope of “g” to “h” is got, such as Eq.4.

(6) Ratio of energy of 0Hz to 5Hz to total energy: “E” is total energy of ECG signal such as Eq.5. “RL” is ratio of energy of 0Hz to 5Hz to total energy such as Eq.6.

$$E = \frac{1}{2\pi} \int_{-\infty}^{\infty} |X(j\omega)|^2 d\omega \quad (5)$$

$$R_L = E_{0-5Hz} / E = \int_0^5 |X(j\omega)|^2 d\omega / \int_0^{\infty} |X(j\omega)|^2 d\omega \quad (6)$$

(7) Location of maximum energy: Maximum energy between 5Hz and 50Hz is got, is shown in figure 4. “a” is the maximum energy.

(8) Ratio of energy of more than 100Hz to total energy.

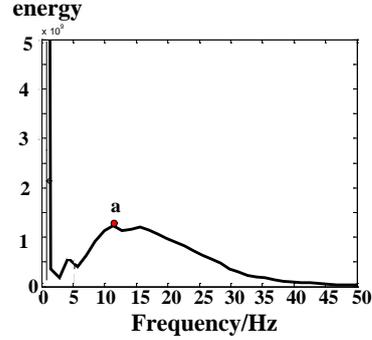


Figure 4 energy spectrum of ECG signal from 0Hz to 50Hz

### 3 Experimental data

Data in this paper is about 5 people from MIT database and 10 people from PTB database, each data contains 100 RR intervals for feature extraction in frequency domain, namely there are 100 feature vectors, with a total number of 1500. Taking 70 data to build a template, namely a total of 1050 data . 30 data is used as test data, namely a database whose total of 450. Using correlation coefficient and neural network to classify respectively .

### 4 Identify

Template library:  $x_n$  denotes the n-th feature data,  $X_{ij}$  denotes the j-th feature vector of the i-th person, shown in Eq.7. Template vector is denoted by matrix  $M_i$ , totally 15 templates, shown in Eq.8.

$$X_{ij} = [x_1 \quad x_2 \quad \cdots \quad x_n] \quad (7)$$

$$M_i = (X_{i1} \quad X_{i2} \quad \cdots \quad X_{ij} \cdots \quad X_{i70})^T, i=1,2,3,\dots,15; j=1,2,3,\dots,70; n=8 \quad (8)$$

### 5 Correlation coefficient

The 20 feature vector is randomly selected from 100 feature vectors for each people, denoted as  $Y_{iz}$ , namely the z-th feature vector of i-th person. Calculate the correlation coefficient of the 70 feature vectors of the template database with the 20 feature vector. We can get 1400 coefficient, then calculate its average, use the number as the coefficient of the template. Show in Eq.9.

$$\rho_j = \frac{Cov(X_{ij}, Y_{iz})}{\sqrt{D(X_{ij})} \sqrt{D(Y_{iz})}} \quad j=1,2,3,\dots,70; z=1,2,3,\dots,20 \quad (9)$$

$D(X_{ij})$  is variance of  $X_{ij}$ ,  $Cov(X_{ij}, Y_{iz})$  is covariance between  $X_{ij}$  and  $Y_{iz}$ . The mean of coefficient is shown in Eq.10:

$$\rho_{\text{mean}} = \sum_{j=1}^N \rho_j / N \quad N = 1400 \quad (10)$$

#### Steps to identify

(1)The correlation coefficient of the template is got and sorted. Sort the correlation

Coefficient got from Eq.9 in descending order, is shown in table 1:

Table 1 the correlation coefficient about 10 templates

| template | correlation coefficient | template | correlation coefficient |
|----------|-------------------------|----------|-------------------------|
| 78       | 0.677260                | 122      | 0.910050                |
| 17       | 0.697759                | 39       | 0.916520                |
| 202      | 0.76333                 | 101      | 0.937656                |
| 20       | 0.794164                | 103      | 0.966630                |
| 28       | 0.819957                | 43       | 0.987152                |

|     |          |    |          |
|-----|----------|----|----------|
| 45  | 0.850485 | 57 | 0.998166 |
| 27  | 0.873535 | 21 | 0.999889 |
| 123 | 0.902674 |    |          |

(2) Calculate the correlation coefficient of test group and template group.

There are 450 data from test group, calculating correlation coefficient between each data and the 70 data of the template group, then calculating the average of 70 correlation coefficient, as the coefficient of the test group with the template. So every test data gets 15 correlation coefficient.

(3) Recognizing

Getting the largest number of the 15 correlation coefficient, comparing it with the sorted template coefficient, For example, if the largest number is 0.97, which is larger than that of template 103, smaller than that of template 43, and the gap between it and 103 is smaller than that of it and 43, so defined it as template 103.

### BP network

BP network structure has three layers: input layer, hidden layer and output layer. The unit number of input and output layer is decided by the number of dimensions and categories of the input vector. In this paper, we got 8 characters of each person which are used as the input data, so the input node is 8. As hidden layer has nodes two times than input layer, so we select 18 nodes, output layer select 15 nodes.

Input vector is shown as Eq.11:

$$X=(x_1, x_2, \dots, x_i, \dots, x_8)^T \quad l=1, 2, 3, \dots, 8 \quad (11)$$

Output vector of the hidden layer is shown as Eq.12 and Eq.13:

$$Y=(y_1, y_2, \dots, y_m)^T \quad m=1, 2, 3, \dots, 18 \quad (12)$$

$$y_j = f\left(\sum_{i=1}^l v_{ij}x_i\right) \quad j=1, 2, \dots, m \quad (13)$$

Output vector of the output layer is shown as Eq.14:

$$O=(o_1, o_2, \dots, o_n)^T \quad n=1, 2, 3, \dots, 15 \quad (14)$$

## 6 Result

The result of human identification by ECG signal based on BP network and correlation coefficient is shown in table 2. FRR (False Reject Rate) is that they should have match each other but they do not. FAR (False Accept Rate) is that they should have not match each other but they do.

The result of human identification by ECG signal based on correlation coefficient shows that the FRR is higher FAR is acceptable. So the method is feasible. The reason of the low recognition rate maybe is that feature selection is inappropriate or small amount of data. Compared with it, the result of human identification by ECG signal based on BP network achieves good results, but there is still room to improve.

Table 2 recognition based on correlation analysis and correlation coefficient

| Test number | FFR of correlation coefficient (%) | FAR of correlation coefficient (%) | FFR of BP network (%) | FAR of BP network (%) | Recognition rate of correlation coefficient (%) | Recognition rate of network (%) |
|-------------|------------------------------------|------------------------------------|-----------------------|-----------------------|---|---------------------------------|
| 17          | 100                                | 5                                  | 0                     | 1.56                  | 50  | 100.0                           |
| 20          | 90                                 | 1                                  | 0                     | 0                     | 10  | 89.3                            |
| 21          | 0                                  | 16                                 | 0                     | 0                     | 100   | 94.4                            |
| 27          | 100                                | 0                                  | 13.3                  | 0                     | 0   | 94.3                            |
| 28          | 100                                | 4                                  | 3.33                  | 0                     | 0   | 100.0                           |
| 39          | 30                                 | 8                                  | 0                     | 0.44                  | 70  | 100.0                           |
| 43          | 100                                | 32                                 | 3.33                  | 0                     | 0   | 96.0                            |
| 45          | 90                                 | 5                                  | 6.67                  | 0                     | 10  | 92.0                            |
| 57          | 60                                 | 11                                 | 0                     | 0.22                  | 60  | 100.0                           |
| 78          | 90                                 | 9                                  | 0                     | 0                     | 10  | 97.1                            |
| 101         | 33                                 | 4                                  | 0                     | 0                     | 33  | 100.0                           |
| 103         | 100                                | 2                                  | 0                     | 0                     | 0   | 100.0                           |
| 122         | 50                                 | 0                                  | 0                     | 0.22                  | 50  | 100.0                           |
| 123         | 60                                 | 6                                  | 13.3                  | 0.67                  | 40  | 86.7                            |
| 202         | 40                                 | 2                                  | 1                     | 0.22                  | 60  | 96.7                            |

## 7 Summary

With the development of science and technology, more people begin to focus on the security and reliability of information. The traditional technology of identity recognition has failed to meet people's need for security now. ECG identification attracts widespread attention with its advantage that the ECG signals are difficult to be replicate, easy to be gotten, etc. The frequency-domain characteristics of the ECG signals are acquired based on the method of Fourier transform of the signals. According to the analysis of the correlation coefficient and neural network analysis respectively, higher recognition rate is achieved by the neural network over the correlation coefficient. Future research will focus on the work below: First, more about other features of the ECG waveform frequency domain will be analyzed and further work about selections of characteristics will be done; Second, the amount of experimental data will be increased to ensure the stability of identification and improve the recognition accuracy.

## Acknowledgements

This work was financially supported by the Tianjin Natural Science Foundation (No.10JCYBJC00700).

## References

- [1]B.P.Simon , C.Eswaran . An EGG classifier designed using modified decision based neural network[J]. Computer and Biomedical Research, 1997, 30 (4) : 257~272.
- [2]Biel L, Petterson O, Philipson L, et al. ECG analysis: a new approach in human identification[J]. Instrumentation and Measurement, IEEE Transactions on, 2001, 50(3): 808-812.
- [3]Tawfik M M, Selim H, Kamal T. Human identification using time normalized QT signal and the QRS complex of the ECG[C]//Communication Systems Networks and Digital Signal Processing (CSNDSP), 2010 7th International Symposium on. IEEE, 2010: 755-759.
- [4]Bhagavatula C, Venugopalan S, Blue R, et al. Biometric identification of cardiosynchronous waveforms utilizing person specific continuous and discrete wavelet transform features[C]//Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE. IEEE, 2012: 4311-4314.
- [5]Neophytou N, Kyriakides A, Pitris C. ECG analysis in the Time-Frequency domain[C]//Bioinformatics & Bioengineering (BIBE), 2012 IEEE 12th International Conference on. IEEE, 2012: 80-84.
- [6]Pal S, Mitra M. Detection of ECG characteristic points using multiresolution wavelet analysis based selective coefficient method[J]. Measurement, 2010, 43(2): 255-261.
- [7]Shen J, Bao S D, Yang L C, et al. The PLR-DTW method for ECG based biometric identification[C]//Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE. IEEE, 2011: 5248-5251.
- [8]Naraghi M E, Shamsollahi M B. ECG based human identification using wavelet distance measurement[C]//Biomedical Engineering and Informatics (BMEI), 2011 4th International Conference on. IEEE, 2011, 2: 717-720.

- [9]Chiu C C, Chuang C M, Hsu C Y. A novel personal identity verification approach using a discrete wavelet transform of the ECG signal[C]//Multimedia and Ubiquitous Engineering, 2008. MUE 2008. International Conference on. IEEE, 2008: 201-206.
- [10] Plataniotis K N, Hatzinakos D, Lee J K M. ECG biometric recognition without fiducial detection[C]//Biometric Consortium Conference, 2006 Biometrics Symposium: Special Session on Research at the. IEEE, 2006: 1-6.
- [11] Chan A D C, Hamdy M M, Badre A, et al. Wavelet distance measure for person identification using electrocardiograms[J]. Instrumentation and Measurement, IEEE Transactions on, 2008, 57(2): 248-253.