Spring 2016	Report#8 - 03/04/2016	Ahmet Can Ozbek

I. Tasks achieved last week

- Obtained classification and EER (Equal Error Rate) results.
- PQRST peak detection of the ECG signal is completed.

II. Feedback and Interaction

- ***Prof. Kuo's F/I***

  o Speaker Identification is similar to our study of ECG biometrics, so those papers should be analyzed further.

III. Report

Improvement of Identification and Authentication from ECG signals

I am using the methods explained in the paper "Dry Contact Fingertip ECG-based Authentication System using Time, Frequency Domain Features and Support Vector Machine" as a starting point for the identification and authentication of individuals from ECG signals. For classification, the feature vector is basically the amplitude values of the ECG signal around the R peaks. The previous results I got was as follows:
*Training Data: Accuracy = **98.7302%** (4976/5040) (classification)
*Testing Data: Accuracy = **90.582%** (856/945) (classification)
*EER(Equal Error Rate) = **3.3862%**

Recently, I investigated on parameters and adjusted them to give an improved result. The method and the new parameters are as follows:

- Dataset and Training Testing Data Partition:
  I am using CYBHi dataset for this study. This dataset has ECG recording of 63 subjects. In each recording of subjects, first 80 heartbeats are used for training data and the next 15 heartbeats are used for testing data.
- Creating train and test data:
  1)First, noisy segments are removed from the ECG signal by the implementation mentioned in the first part of this report.
  2)ECG signal is filtered by a bandpass filter with cutoff frequncies 0.45Hz and 35Hz with an FIR filter with order 75
  3)R peaks are detected by Pan Tompkin's algorithm.

4)For each heartbeat, 275 samples before and 265 samples after the R peak is cropped to be used as a temporal feature vector (R-275, R+265)
5)For each heartbeat, 80 samples before and 80 samples after the R peak is cropped and taken 45 point FFT to be used as a frequency component feature vector.
6)Temporal feature vector and frequency component feature vector is cascaded to make one feature vector for each heartbeat.
(temporal feature vector + frequency component feature vector = Dimension of the feature vector)
$(275 + 265 + 1) + (45) = 586$
As a result dimensionality of the feature vector for each heartbeat is 586.
7)The training and testing data is normalized between 0 and 1 for SVM classification.

- SVM classifier and results:
  By using SVM with RBF Kernel with parameters: gamma = 0.02 and C = 1000, the results on training and testing data is as follows:
  *Training Data: Accuracy = **99.246%** (5002/5040) (classification)
  *Testing Data: Accuracy = **91.6402%** (866/945) (classification)
  *EER(Equal Error Rate) = **2.9333%**

As we can see the results have improved compared to the previous results. I would also like to note that, if we take the majority classification result vote of 15 heartbeats of an individual, we get **100%** accuracy on identification performance.

In figure 1, we see a summary of EER(Equal Error Rate) results in existing literature. As in my implementation training data and testing data come from the same session (within-session), I would like to share some EER results obtained from within-session analysis in the existing literature.

IV. Plan for the next week

- Try to get results similar to the papers using feature extraction techniques
- Start on deep learning

V. Milestone

- PQRST peak detection of the ECG signals in CYBHi data is completed.
- A significant accuracy for identification from ECG signal is obtained by implementation.

TABLE II
AUTHENTICATION PERFORMANCE FOR WITHIN-SESSION ANALYSIS

Researchers	Equal Error Rates (%)				
	Literature	Train 8, Test 8	Train 16, Test 16	Train 32, Test 32	Train 64, Test 64
Agrafioti *et al.* [17]	0.6	3.88	0.85	0.57	0.38
Chan *et al.* [12]	-	5.82	3.84	3.02	2.26
Chiu *et al.* [40]	0.83 - 0.86	4.15	2.64	1.76	1.01
Coutinho *et al.* [53]	-	42.54	38.92	35.14	33.34
Fang and Chan (SC) [33]	-	19.81	19.1	19.03	18.82
Fatemian and Hatzinakos [41]	-	8.69	5.99	4.37	2.26
Irvine *et al.* [43]	-	2.25	1.74	1.26	0.69
Khalil and Sufi [44]	-	5.28	2.64	1.56	1.13
Li and Narayanan (HPE+SVM) [18]	0.55	2.19	1.24	1.17	0.96
Lourenço *et al.* [58]	13	12.01	9.3	6.56	5.25
Molina *et al.* [39]	2	19.99	16.31	16.27	15.98
Molina (M) [39]	-	13.71	7.53	6.16	5.57
Odinaka *et al.* (FS) [11]	0.02	1.89	0.93	0.38	0.03
Odinaka *et al.* (NFS)[11]	-	1.93	1.04	0.51	0.06
Sufi *et al.* [47]	-	27.39	21.97	17.14	13.42
Wan and Yao *et al.* [27]	-	7.98	2.15	0.75	0.27
Wang *et al.* (DCT) [25]	-	3.9	2.22	1.74	1.36
Wübbeler *et al.* [14]	-	1.08	0.57	0.57	0.38
Yao and Wan [48]	-	24.46	22.32	20.63	18.49
Ye *et al.* [32]	-	5.11	2.84	1.64	1.13

"Train 8, Test 8" represents training on 8 heartbeats (or 8 s) and testing on 8 heartbeats (or 8 s) from the same session. "Molina (M)" represents a modified Molina algorithm. DCT = Discrete Cosine Transform; FS = Feature Selection; HPE = Hermite Polynomial Expansion; NFS = No Feature Selection; SC = Spatial Correlation; SVM = Support Vector Machine.

Figure 1: Some Results of EER(Equal Error Rate in existing literature). From the paper *ECG Biometric Recognition: A comparative Analysis, Ikenna Odinaka, Student Member, IEEE, Po-Hsiang Lai, Student Member, IEEE, Alan D. Kaplan, Member, IEEE, Joseph A. O?Sullivan, Fellow, IEEE, Erik J. Sirevaag, and John W. Rohrbaugh