An Exercise ECG Database With Synchronized Exercise Information

Yuanjing Yang[1], Lianying Ji[1], JiankangWu[1]
[1] Sensor Networks and Applications Joint Research Center (SNARC)
Graduate University and Institute of Automation, Chinese Academy of Sciences, Beijing
E-mail: yangyuanjing09@mails.gucas.ac.cn TEL: +86-10-88256015

***Abstract-* Cardiovascular system is affected not only by diseases but also other factors among which the activity is the major one. Currently, there is a lack of ECG database with precise activity information recorded for research. In this paper, we present a database which includes both ECG data and activity signals. The data are collected in both strictly controlled multi-activity sessions and free daily life activities, and include synchronized 3-lead ECG data and 3-axis acceleration. The data collection scenario, storage format and supporting software are provided. Easy-to-use data reading, activity processing software is provided. The database is supposed to be used for activity-related cardiovascular system monitoring, disease diagnosis and related research.**

I. Introduction

Open physiological signal database is one of the most important support conditions for world-wide medical, biological, sport and information processing researches. A lot of databases with well-controlled collection process, professional annotation, and support software have been opened. The types of data are becoming richer and richer with advances of the research interests of medical and other professionals.

PhysioBank built by Massachusetts Institute of Technology (MIT) is a large and growing archive of well-characterized digital recordings of physiologic signals and related data for use by the biomedical research community and now contains over 50 databases that may be freely downloaded. Recently, it is internationally acknowledged that there are only three ECG databases can be used as a standard, namely, MIT-BIH database, AHA database and ST-T database. Among them, MIT-BIH database provided by MIT is used more widely. In PhysioBank, the long-term ST (LTST) database built by F.Jager and his fellows has become a research resource for algorithm development and physiologic studies of transient myocardial ischemia [1]. The LTST database contains 24-hours ambulatory records selected from Holter recordings obtained in routine clinical practice settings in Europe and in the United States. Each record includes a header file, a signal file, QRS and ST reference annotation files, a reference annotation file with measurements obtained on average heart beats and attached back to individual heart beats, and a clinical summary [2]. Besides, Pablo Laguna and his fellows built a QT database for evaluation of algorithms for measurement of QT and other waveform intervals in the ECG [3]. The QT Database includes ECGs which were chosen to represent a wide variety of QRS and ST-T morphologies, in order to challenge QT detection algorithms with real-world variability. The records were chosen primarily from existing ECG databases, including the MIT-BIH Arrhythmia Database, the European Society of Cardiology ST-T Database and several other ECG databases collected at Boston's Beth Israel Deaconess Medical Center. The records are provided in the MIT-BIH Database format. Each record includes a signal file, a header file, and several annotation files.

It is well-known that cardiovascular system is heavily affected by multiple factors such as sex, age, emotion and activity. Among all these factors, activity is the most important one. Although exercise ECG records can be found in current databases, there is a lack of precise exercise information. The description of the scenario is obscure, and the synchronized exercise type and intensity information are not recorded. This information is critical for researchers to evaluate the state of the cardiovascular system. Since 2005, Sensor Networks and Applications Research Center (SNARC) has been doing its research on context-aware healthcare. The lack of such an ECG-plus-exercise database is noted, and experimental activity-aware ECG data collection is carried out. The data collected are built as a database and are currently open on our website.

The database contains both 3-lead ECG data and 3-axis acceleration data. The short-term records include well-controlled exercise type and intensity. The long-term 24-hour data record free daily life activities. The data format, experimental activity session and supporting software are also introduced. The database is supposed to be used for exercise-related ECG processing and cardiovascular state evaluation researches.

The rest of this paper is organized as follows: The experimental design and data collection processes are introduced in Section II. Section III presents the format of header, data, and annotation files. Supporting software is introduced in Section IV. Section V provides the conclusion.

II. The Data Collection Experiments

A. The Device For Data Collection

The data contained in the database were all acquired using our own context-aware heart monitor called uCare which is shown in Fig.1. It is a Holter-like wearable heart monitor device, which can 1) record 3-lead ECG signals, 2) record 3-axis acceleration signals, 3) record real-time clock, 4) and store them into a microSD card.

The device can be worn by the subjects around the chest. The device can work for 48 hours continuously when fully charged. The sampling rate of ECG is 300Hz, while the sampling rate of acceleration signal is 100Hz. Both ECG and acceleration are sampled using 8-bit resolution.

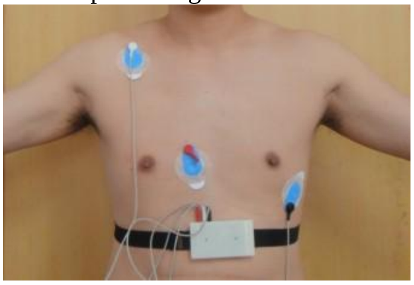

Fig. 1 uCare and its wearing

B. Subjects in the Expeirment

20 young healthy men aging from 23-30 participated in the experiments. The subjects were 13 men and 7 women, with body mass index (BMI) from 18-27, and with no history or clinical signs of cardiovascular or pulmonary disease. Candidates were not taking any kind of medication and did not show abnormal blood pressure or ECG patterns.

C. The Data Collection Experiments

The data were collected from three experiments with different activity contexts. The first one is a short term session including consecutive static and middle intensity exercise. The second one is 24-hour containing free daily life activities. The third one includes multiple static and exercise phases. The details of the three experiments are as follows.

a) The two-phase short-term exercise ECG

This experiment included a two-phase activity, i.e., consecutive sitting and running sessions. The data were collected for researches on comparison or transition between ECG signals and states of cardiovascular system at rest and during exercise.

This two-phase experiment was conducted in working conditions, which were not strictly controlled. The resting session was carried out indoor and the running session was outdoors. Ten of the above-mentioned subjects participated in this experiment. The subjects were asked to sit for 30minutes with body relaxed. During sitting, they were free to do small movements lasting for short time, for example, programming, web surfing, watching movies. Immediately after the resting session, the subjects were asked to do the jogging with the device still working. The intensity was not rigorously controlled, while most of the subjects running with middle or light intensity, for example, 5 miles/hour. Both sitting and jogging were lasting for about 30 minutes. By the end of the running session, they were asked to sit down for a rest for about 15 minutes to recover the heart rate. The device recorded all above-mentioned sessions and was shut down after that.

b) The collection of ECG data with various intensity exercises

The experiment included multiple consecutive resting and running sessions with controlled stable intensity. The data collected in this experiment were supposed to be used for researches on comparison between ECG waveforms and cardiovascular state between multiple intensity levels. For example, one can obtain ECG morphological changes under different exercise intensities, which can be seen as dynamics of the morphological changes. Meanwhile, the state during the recovery phase after each exercise under different intensity is also valuable.

17 of the subjects participated in this experiment. The experiment was conducted in gym with controlled conditions. This experiment was conducted on the treadmill in the gym with temperature 20^{o}C. Besides, before and during the experiment, the subjects didn't drink water.

The subjects were asked to select three different speeds according to their own physical qualities. The three speeds were supposed to be the light, middle and intense intensity for the each subject respectively. While most of male subjects chose 5miles/hour, 7 miles/hour, and 9 miles/hour, most of female subjects chose 3miles/hour, 5miles/hour, and 7miles/hour. Each of the running session with one chosen speed lasted for 10 minutes following with 15 minutes resting. During the resting session, the subjects were almost totally static without any significant movements. Besides, they were also asked to have a rest for a quarter before and after the whole procedure respectively.

c) The collection of 24-hour ECG data

This experiment included 24-hour daily life activities of each participated subjects. The data collected in this experiment has two usages at least. One is for researches on cardiovascular state of daily life activities, which is usually different from exercise in the gym. Daily life activities have more series of short term moments and are more random. Another is for researches on cardiovascular state during different biological clock stages. For example, the wake-sleep cycles have significant impact on heart and its regulation state.

All the above-mentioned 20 subjects participated in this experiment. The procedure lasted for 24 hours, while the starting time was freely chosen. During the time, the subjects were asked to do their daily life activities, but normal sleep-wake time was followed. Because all the subjects are graduate students, their daily life are similar. The daily activities include morning washings, eating, working in the lab, and sleeping. Two exercise sessions lasting for 10 minutes is also asked conducted during the procedure, but can be conducted at any time. The activities were also recorded by the subjects manually, which gave each activity a recorded types and corresponding starting and ending time. Meanwhile, the ECG signals and acceleration signals were recorded by our uCare devices.

III. The File Format

The data collected in above mentioned three experiments are further re-formatted, and some assistive files are provided to help understand and process the data. Thus, each data record includes a header file, a data file, and an annotation file. The header file details the name and attributes of its associated data file. The data file stores the ECG and acceleration data. The annotation file records the activity sessions.

A. The Header File

The contents of header file consist of one or several lines of ASCII, and at least contain a recording line. Usually, there are also signal specification lines, and information annotation line. The recording line records the name of the signal, the signal number, the time and the data of the start of sampling. The signal specification line mainly contains the file name storing the signal, storage format, ADC resolution, ADC zero value, sampling initial value, sampling rate, sampling number of each signal, and so on. They are separated by one or several space. The information annotation line usually locates in the end of the file. It starts with "#" and the contents are generally a simple description of the subject. One example of the header file is as follows:

```
yyj 5 18:01 2011/4/11
yyj.bin 80 8 128 134 300 1608450 ECG MLII
yyj.bin 80 8 128 220 300 1608450 ECG V5
yyj.bin 80 8 128 163 100 536150 ACC x
yyj.bin 80 8 128 119 100 536150 ACC y
yyj.bin 80 8 128 185 100 536150 ACC z
#age: 25 sex: F
#none
```

The first line of this header file is the recording line, which specifies that the recording includes five signals whose sampling time and data are also recorded. Five signal specification lines are followed. It is clearly seen that five signals are all included in the file yyj.bin and stored in format 80. Their ADC resolution is eight bits, and zero values are all 128. Following are their sampling initial values. The first two signals are ECG signals. The sampling rate of ECG signals is 300 samples per second (SPS). The length of each signal is 1608450 sampling points. It is noted that the two signals are derived from MLII and V5 leads. The other three signals are acceleration signals. The sampling rate is 100 SPS. The length of each signal is 536150 sampling points. In the end of the file is the information annotation line. It mainly notes the age and sex of the subject. The letter "F" stands for "female" while "M" stands for "male". The word "none" denotes the subject didn't take any drug during the data collected procedure.

B. The Data File

To make our database easy to use, popular ECG data storage format is followed. There are eight data storage formats which are used in MIT-BIH database for ECG and other physiological data. According to the bit resolution of the data, the data can be stored using format 8, format 16, format 80, format 212, and format 310, etc. Format 80 is chosen to represent the collected data, which is most suitable for our 8-bit resolution data. By format 80, one data sample is represented by one byte offset binaries. One particular side of our data is that, each data record contains five signals, two of which are ECG signals (ECG1 and ECG2) and the other three are acceleration signals (ACC_X, ACC_Y and ACC_Z). Furthermore, the sampling rate of ECG signals is 300 SPS while the sampling rate of acceleration signals is 100 SPS. Therefore, The ECG and acceleration signals are stored in the way shown in Table I.

TABLE I. FORMAT OF THE DATA FILE

ECG1			ECG2			ACC_X	ACC_Y	ACC_Z
S1	S2	S3	S1	S2	S3	S1	S1	S1

[a]S1, S2, S3 mean the first, the second and the third sampling.

C. The Annotation File

Each record has an annotation file which explains the activity sessions. The annotation file is an ASCII-coding text file and has multiple lines. Each line describes one session of an activity during the experiment. All the sessions are stored sequentially according to the time. In every line, there are three columns which are starting sampling, ending sampling, and activity types. The time is represented in ECG sample indexes to synchronize ECG data easily. The activity types are represented by numbers shown in Table II.

TABLE II.
THE ACTIVITY TYPES AND THEIR NUMBERING

Activity type	number
Sit	1
Stand	2
Lay	3
Walking	4
Running with low intensity	5
Running with middle intensity	6
Running with high intensity	7
Random	8

One example of the annotation file is as follows:

```
1          200000     #1
200001     415000     #5
415001     643000     #1
643001     838500     #6
838501     1105800    #1
1105801    1300000    #7
1300001    1624950    #1
```

IV. The Supporting Software

To facilitate the usage of the presented database, we also provide software help process the data and extract information. Two software files are provided. One is for reading data according to the data format. Another is for extracting the activity information from the acceleration signals. All the software is written in m file which can be run with the Matlab software.

A. The Software for Reading Data

This software is provided to read and retrieve the data stored in the database. To retrieve the data and information, all above-mentioned files may be used. The software takes following steps: 1) Read the head files to obtain necessary information such as sampling rate, starting real time, number of samples and binary offset in the file. 2) Reading data from the data file according to the format 80 and number of samples. 3) Process the ECG and acceleration data, for example, retrieving the original amplitude and sampling time. 4) Read the annotation file to obtain the activity information. 5) Demonstrate the data in figures. The result of data reading software is shown in Fig.3.

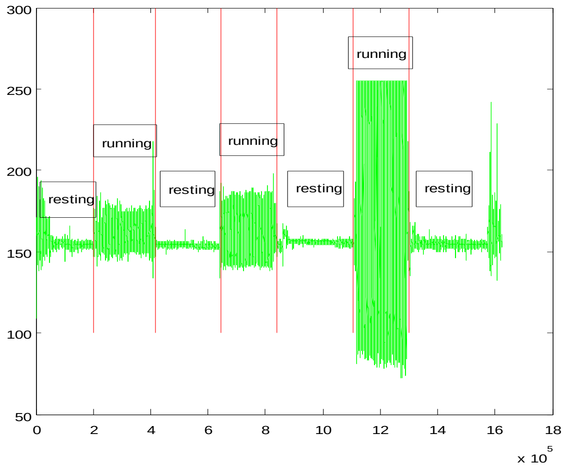

Fig. 3 The result of data reading program

B. The Activity Processing Software

A software which does activity classification and intensity computing using acceleration signals is also provided. The software reads the data file directly and extracts 3-axis acceleration signals as input. The processing is based on the method in [4]

The activity classification is based on a decision tree. Using single 3-axis accelerometer, we could recognize following activity types: walking, running. The postures that can be classified includes standing, sitting, lying flat, lying on left side, lying on right side. Static postures and dynamic activities are first classified by examining existence of the variation of acceleration signals. Then, subdivision is made to classify different postures according to proportions of 3-axis acceleration signals, which are projected components of gravity signal on the three axes. Different dynamic activities are classified using the different patterns of acceleration signals.

The activity intensity has linear relationship with total fluctuation area (TFA) of acceleration signals [5]. Therefore TFA is used to indicate intensity level. Higher the TFA value means higher the activity intensity. Defining $a_{k,i}$ as acceleration signal of axis k at time i, and N as length of a analysis time window, TFA is defined by

$$\begin{cases} FA_k = \sum_{i=1}^{N} \left| a_{k,i} - \overline{a}_k \right|, k = \{1,2,3\} \\ TFA = \sum_{k=1}^{3} FA_k \end{cases}$$

The step frequency of walking or running can also be calculated by computing the cycles of fluctuation of acceleration signals. The acceleration axis which is vertical to trunk has the biggest fluctuation amplitude, and thus is used. Moving average with 32 points window is applied first. Intervals of peaks of the sinusoidal-like waveform are used to calculate step frequency according to

$$SF = \left(\frac{1}{\sum (1/T_p(k))} \right)$$

where T_P is interval of peaks, and T_w is length of the window

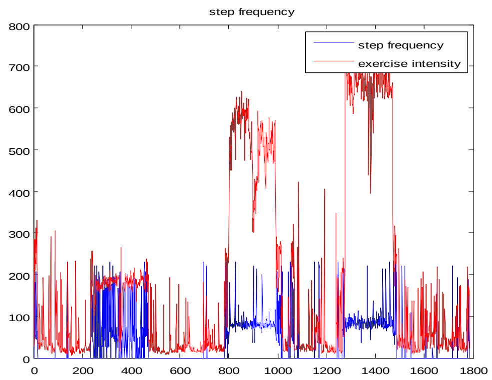

Fig. 4 The exercise intensity and step frequency calculation result

V. Conclusion

In this paper, an ECG database which includes synchronized activity information is presented. The database includes both short-term records with controlled activity types and intensity, and long-term records with normal daily life activity. The database can be used for activity-related ECG processing such as exercise-ECG denoising with intensity as reference. It can also be applied for exercise-related cardiovascular state estimation and evaluation. One of such applications is context-aware HRV analysis, which can not only obtain more precise HRV at rest, but also HRV during different activity intensity. The idea has been verified by our previous research. Exercise-driven heart rate and ECG morphological changes are another potential application. Consistent with the complex dynamic nature of the cardiovascular system, the database will promote researches on the context-aware heart monitoring and diagnosis. All the records and supporting software can be freely downloaded on our website, and we will submit them to PhysioBank in near future.

References

[1] Franc Jager, Alessandro Taddei, Michele Emdin, Gorazd Antolic, Roman Dorn, George B. Moody, Boris Glavic, Ales Smrdel, M Varanini, Mitja Zabukovec, Simone Bordigiago, Carlo Marchesi, and Roger G. Mark. The Long-Term ST Database: A Research Resource for Algorithm Development and Physiologic Studies of Transient Myocardial Ischemia. *Computers in Cardiology 2000*, pp. 841-844.
[2] Franc Jager, George B. Moody, Alessandro Taddei, Gorazd Antolic, Mitja Zabukovec, Maja Skrjanc, Michele Emdin, and Roger G. Mark. Development of a Long-Term Database for Assessing the Performance of Transient Ischemia Detectors. *Computers in Cardiology 1996*, pp. 481-484, IEEE Press. ISSN 0276-6547.
[3] Pablo Laguna, Roger G. Mark, Ary Goldberger and George B. Moody. A Database for Evaluation of Algorithms for Measurement of QT and Other Waveform Intervals in the ECG. *Computers in Cardiology 1997*, pp. 673-676
[4] Aiguang Li, Lianying Ji, Shaofeng Wang, Jiankang Wu, "Physical Activity Classification Using a Single Triaxial Accelerometer Based on HMM", IET International Conference on Wireless Sensor Network (IET-WSN), pp. 155-160, 2010.
[5] Midorikawa, T., S. Tanaka, et al. (2007). "Evaluation of Low-Intensity Physical Activity by Triaxial Accelerometry." Obesity 15(12): 3031-3038.