Abstract

Signal quality is important to produce reliable results in computer-based analysis of heart sound signal. Automated signal quality assessment becomes a necessary preprocessing step before upcoming analysis. For this purpose, the contributors collected 7893 recordings from open public heart sound databases and did manual annotation for each recording as gold-standard label. These recordings were sampled by different devices, different sensors and in different environments. From point view of patter recognition, it is possible set up a network for automated quality assessment by training. The contributor wish to share the data in patter recognition community for those who are interested in this topic.

Method

This database were collected from four sources. They are listed in the follows.

(1)The training subset of heart sound database for CinC Challenge 2016 The training subset is freely available from the Physionet website consisting of 3153 recordings.

(2) Heart sound signals from “Pascal Classifying Heart Sound Challenge”, named PASCAL in short. These signals was set up in 2012 for the objective of classification too. The data were gathered from two sources. One was from iPhone app and another was from a clinic trial in hospital using a digital stethoscope. There are 859 recordings available.

(3) Heart Sounds Catania 2011, named CTHS in short. This was a collection of heart sounds used for the purpose of biometry by the University of Catania, Italy. It contained heart sounds acquired from 206 people using a digital stethoscope. There are 412 recordings available.

(4) Heart sound data of pulmonary hypertension, named PHHS in short. These data were acquired from 76 patients bedside by the authors’ group from the second attached hospital of Dalian Medical University since 2015. Each signal is non-overlap separated into subsequence while a subsequence is no longer than 6s in length. The operation produces 3875 recordings.

The four databases provide 8299 recordings available. However, to ensure the signal quality can be reliably assessed, those recordings with time length less than 6 s are excluded. Finally, 7893 recordings are remained for signal quality assessment.

In order to develop an automatic signal quality classification algorithm, gold standard annotations for the signal quality of each recording are needed. These gold annotations were obtained from manual annotations by one skilled physician and a senior researcher in the field of heart sound signal processing with 10 years’ experience. Each annotator did these annotations in quiet environments using both headphone listening and visual examination. Each recording was assigned a quality label rating of one to five for the entire recording according to the label scheme.

Table 1 Labelling scheme for heart sound signal quality annotations

|  |  |
| --- | --- |
| Quality label | Quality description |
| 1 | Very bad quality, no heart sound can be heard and identified by human eyes. only noise or only harmonic signal |
| 2 | Bad quality, mostly noise but some heart sounds can be heard and identified by human eyes |
| 3 | Borderline quality, very weak heart sounds but beating rhythms can be recognized, fairly difficult to interpret |
| 4 | Good quality, heart sounds can be easily heard and interpretable, but some noise presents |
| 5 | Excellent quality, almost no noise, heart sounds can be clearly heard, identified by visual check and interpretable with confidence |

The 7893 recordings are saved as “\*.wav”. The quality labels are given in an excel file. The features and the method used for this feature extraction, as well as the quality classification method, can be found at

<https://github.com/tanghongdlut/signal-quality-assessment-of-heart-sound-signal>

Contact: Hong Tang, [tanghong@dlut.edu.cn](mailto:tanghong@dlut.edu.cn),

School of Biomedical Engineering, Dalian University of Technology