Classification of Doppler Ultrasound Signal Quality for the Application of Fetal Valve Motion Identification

Faezeh Marzbanrad[1], Yoshitaka Kimura[2], Miyuki Endo[2], Marimuthu Palaniswami[1], Ahsan H Khandoker[1,3]

[1] The University of Melbourne, Parkville, Australia
[2] Tohoku University, Sendai, Japan
[3] Khalifa University of Science, Technology and Research, Abu Dhabi, UAE

Abstract

One dimensional Doppler Ultrasound (DUS) is a commonly applied technique for fetal heart rate monitoring, but it can also be used to identify the timings of fetal cardiac valve motion. These timings are required to estimate the fetal cardiac intervals, which are fundamental and clinically significant markers of fetal development and well-being. Several methods have been proposed in previous studies to automatically identify the valve movement timings using 1-D DUS and fetal Electrocardiography (fECG) as a reference. However DUS is highly susceptible to noise and variable on a beat-to-beat basis. Therefore it is crucial to assess the signal quality to ensure its validity for a reliable estimation of the valve movement timings. An automated quality assessment can provide the operator with an online feedback on the quality of DUS during data collection. This paper investigates automated classification of the DUS signal quality using Naive Bayes (NB) classifier. The quality of 345 beats of DUS signals collected from 57 fetuses was assessed by four independent annotators and used for training and validation of the classifier. Using Fleiss kappa test, a fair agreement was found between the raters with overall $\kappa = 0.3$. The performance of the classification was tested by 10-fold cross validation. Results showed an average classification accuracy of 86% on training and 84% on test data.

1. Introduction

Fetal cardiac intervals have been clinically used to characterize the fetal cardiac function and they provide more information on fetal well being than heart rate alone. In particular, Systolic Time Intervals (STI) have been established as indicators of myocardial function and fetal development [1–3]. Cardiac intervals are estimated from the onset of QRS complex of fetal Electrocardiography (fECG) together with the opening and closing time of heart valves. Early studies in the 1980s proposed non-invasive methods to estimate the intervals using the one dimensional Doppler Ultrasound (DUS) signal and fECG [4–7]. These studies used band pass filtering approaches to extract the high frequency component of the DUS, from which the valve movements were identified manually by experts using fECG as a reference. Improved signal processing techniques were later used to separate the component linked to the valve movements more effectively, such as Short Time Fourier Transform (STFT) analysis, multi-resolution wavelet [8–10].

Several automated techniques for identification of valve movements were proposed in our previous studies to overcome the shortcomings of manual methods including their time consuming process and vulnerability to inter and intra observer errors [11–13]. However the pattern and the quality of the DUS signal were found to be variable, even on a beat-to-beat basis [12]. The signal is highly contaminated by noise and its extensive variability and non-stationary characteristics complicate the valve identification. Therefore, an automated DUS quality assessment is required for a reliable estimation of the valve timings and also providing a real time feedback to the operator during data collection. The importance of DUS signal quality assessment for its classic application in Fetal Heart Rate (FHR) monitoring, was investigated in previous studies [14, 15]. This paper focuses on the signal quality assessment for the extended application of DUS signal in valve motion identification.

2. Methods

2.1. Data

One dimentional DUS data were obtained using ultrasonic transducer 5700 (Corometrics Medical Systems, Inc., model: 116) with 1.15 MHz signals. Data were collected from 57 pregnant women at the gestational age of

16 to 41 weeks with healthy single pregnancies at Tohoku University Hospital in Japan. Abdominal ECG signals were also recorded using 12 electrodes (10 electrodes on the mothers abdomen, one maternal reference electrode at the right thoracic position and one on the back). A multi-channel data acquisition system was used to collect the simultaneous DUS and ECG data. All recordings were one minute in length and sampled at 1 kHz with 16-bit resolution. The study protocol was approved by Tohoku University Institutional Review Board and written informed consent was obtained from all participants.

2.2. fECG extraction

Fetal ECG was used as a reference for segmentation of the DUS signal into cardiac cycles. It was extracted from the abdominal ECG recordings using a combination of maternal ECG cancelation and blind source separation with the reference signal (BSSR) [16]. The R-peaks of fECG were then detected automatically applying a lower threshold (e.g. 5 times the mean of fECG over 10 second intervals) and peak detection based on the derivative of the signal.

2.3. DUS signal decomposition

The DUS signal contained components associated with the motion of fetal cardiac valves and walls or other maternal and fetal organs. To isolate the high frequency component of the DUS signal linked to the valve movement, the DUS signal was decomposed by the multiresolution Wavelet analysis, the same as in our previous study [10, 12]. Using a second order complex Gaussian as mother wavelet, the detailed signal of the DUS signal at level 2 (100-200 Hz) was obtained as the valve motion related component. The envelope of the absolute value of this signal was then estimated by interpolating the maxima and smoothing by a low pass filter. Each envelope was segmented into cardiac cycles using the corresponding R-R intervals estimated from fECG. The signal segments were then normalized by subtracting the mean and dividing by the standard deviation. Considering that the valve motion events mostly happen within 350 msec following the R-peak [10–12], this section of the DUS segments was used for quality assessment.

2.4. Signal quality annotation

Signal quality annotation was performed in two phases, using 345 DUS segments. In the first phase, five beats with the closest heart rate to the median of FHR were selected from each recording for training. Total of 285 DUS segments were presented to two medical doctors and two researchers to rate the quality independently. The scoring was based on observing the data to identify four peaks, corresponding to mitral closing (Mc), aorta opening (Ao), aorta closing (Ac) and mitral opening (Mo). Five quality levels were defined as described in table 1 and given to the annotators as instructions on quality rating. Examples of a very good and a very bad quality signal scored by the annotators are shown in figure 1. The possible ranges of Mc, Ao, Ac, and Mo events were shaded with yellow, green, magenta and cyan colors respectively, as guides for the annotators.

Inter-rater agreement was tested by Fleiss kappa test [17, 18]. It calculates the degree of agreement in classification against the completely random rating. Scores of 1 to 5 were assigned to very bad to very good labels. The signals with the average score of below 2.5 and above 3.5 were labeled as unacceptable (60 signals) and acceptable (121 signals), respectively; while others were labeled as ambiguous (104 signals). In the second phase in order to balance the classes, 60 additional poor quality DUS segments as confirmed by the annotators were selected from the recordings and labeled as unacceptable.

Table 1. Description of the the quality levels used for annotation

Quality Level	Quality Description
very good	Mc, Ao, Ac, Mo peaks are clearly detectable with no doubt.
good	Although the signal is slightly noisy, at least 3 events can be clearly detected.
borderline	It is difficult to detect the events, but some traces are observed, or at least two events can be detected.
bad	There is mostly noise, it is impossible to detect the events.
very bad	No trace of the events, only noise.

2.5. Signal quality indices

Twelve features were selected mostly based on the signal properties in the valve motion ranges compared to the remaining time intervals. The plausible valve motion ranges were defined as: Mc:[9-44], Ao:[45-90], Ac:[200-260], Mo:[265-326], all in msec following the segment onset (preceding R-peak) [10, 12]. The features were as follows and all normalized:

- The ratio of the power (SQI_1), number of peaks (SQI_2), mean peak amplitude (SQI_3) and variance (SQI_4) in the valve motion range to the values in the remaining time intervals.
- kurtosis (SQI_5), skewness (SQI_6), Hjorth (SQI_7) parameters and sample entropy (SQI_8: $m = 1, r = 0.1$,

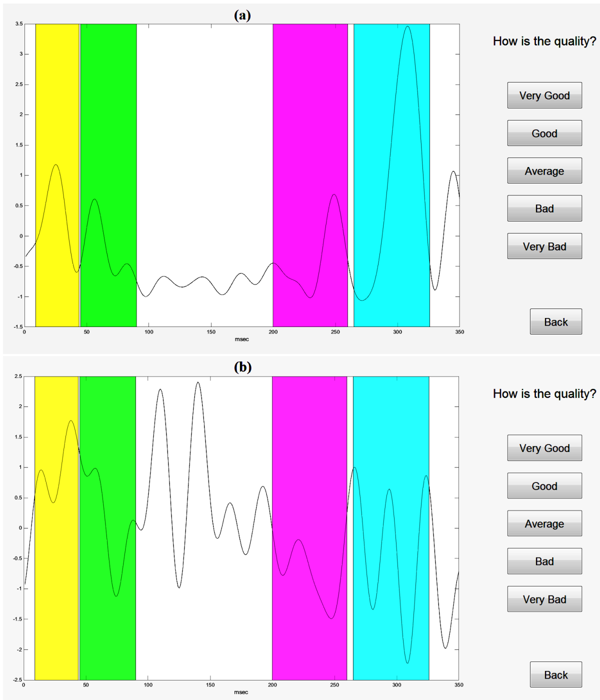

Figure 1. Two examples of annotated signals as very good (a) and very bad (b). The possible ranges of Mc, Ao, Ac, and Mo events were shaded with yellow, green, magenta and cyan colors respectively, as guides for the annotators.

SQI_9: $m = 1, r = 0.2$, SQI_{10}: $m = 2, r = 0.1$, SQI_{11}: $m = 2, r = 0.2$) as described in previous studies [19, 20].

- Minimum ratio of the 2nd to 1st singular value (SQ_{12}) from Singular Value Decomposition (SVD) of a matrix containing consecutive windows of the signal with various sizes: 10, 15, 20,...,100 [19, 21].

2.6. Classification

An overall quality metric was obtained from the quality features $SQI_{1,2,..,12}$. A naive Bayes (NB) classifier with kernel density estimate was used for this purpose. NB classifier is a widely used supervised learning method which is fast and simple to implement [22, 23]. It uses the training data to estimate the conditional distribution of the features given the classes and also distribution of the classes. Then it assumes conditional independence of the features given the classes which dramatically simplifies the estimation of the probabilities. Then it estimates the posterior probability through the Bayes rule and classifies a sample to the most probable class. It is important to note that in practice the features may not be independent while NB still works properly. Since some features did not have normal distribution, kernel density estimate was performed based on the training data [23].

10-fold cross validation was used to evaluate the classification performance, and the accuracy, sensitivity and specificity in train and test sets were calculated.

3. Results

Inter-rater agreement results of Fleiss kappa test showed a fair agreement with overall $\kappa = 0.300$, C.I. (95%) of $\kappa = [0.293 - 0.307]$, and $p < 0.0001$ confirming that the observed agreement was not accidental. Kappa values for the score 1 to 5 were: 0.224, 0.257, 0.232, 0.277, 0.507, respectively.

Sensitivity (Se), was measured as the proportion of unacceptable signals that were correctly identified as unacceptable. Specificity (Sp), was also calculated as the proportion of acceptable signals that were correctly classified as acceptable. Finally, Accuracy (Ac) was measured as the proportion of correctly classified quality of the signals. Results are summarized in table 2.

Table 2. Average classification results (mean ± standard deviation) for the train and test data, based on 10-fold cross validation.

	Accuracy	Sensitivity	Specificity
Train	0.863±0.007	0.832±0.016	0.894±0.013
Test	0.842±0.038	0.800±0.070	0.884±0.059

4. Discussion and conclusion

The quality of the DUS signal is usually affected by noise and also depends on the fetus-transducer orientation. Although the DUS quality assessment has been previously investigated, it was only targeted for improving FHR monitoring [14, 15]. Results of our study show that the DUS quality can also be assessed in more detail, based on its reliability for valve motion identification.

A real time feedback on the signal quality during data collection would improve the quality of DUS signal for a more accurate estimation of fetal cardiac intervals. Results show that the NB classifier can be used for an accurate classification of the signal quality. NB also requires a short computational time, can be simply implemented and is not sensitive to irrelevant features. However further investigation of other classification techniques are required particularly to improve the sensitivity, in order to provide a reliable feedback to recollect or exclude the poor quality data for further analysis. The classification performance can also be improved by investigating better discriminative features in future studies.

A limitation of the proposed method is the dependance of features on the predefined range of the valve motions. Although the ranges were assumed wide enough to accommodate the variation of the intervals with age or heart rate, the validity of the measures should be assessed for abnormal cases in future studies.

Acknowledgements

This study was supported by an Australian Research Council Linkage grant (LP100200184) between the University of Melbourne, Tohoku University and Atom Medical Corporation in Japan.

References

[1] Zador IE, Wolfson RN, Pillay SK, Timor-Tritsch IE, Hertz RH. Fetal cardiac time intervals and their potential clinical applications. Clinical Obstetrics and Gynecology 1979; 22(3):651–663.

[2] Murata Y, Chester B Martin J. Systolic time intervals of the fetal cardiac cycle. Obstetrics Gynecology 1974; 44(2):224–232.

[3] Wolfson R, Zador I, Pillay S, Timor-Tritsch I, Hertz R. Antenatal investigation of human fetal systolic time intervals. American Journal of Obstetrics and Gynecology 1977; 129(2):203.

[4] Murata Y, Martin CB, Ikenoue T, Lu P. Antepartum evaluation of the pre-ejection period of the fetal cardiac cycle. American Journal of Obstetrics and Gynecology 1978; 132:278–284.

[5] Sampson MB. Antepartum measurement of the preejection period in high-risk pregnancy. Obstetrics Gynecology 1980;56(3):289–290.

[6] Organ L, Bernstein A, Hawrylyshyn P. The pre-ejection period as an antepartum indicator of fetal well-being. American Journal of Obstetrics and Gynecology 1980; 137(7):810–819.

[7] Koga T, Athayde N, Trudinger B, Nakano H. A new and simple doppler method for measurement of fetal cardiac isovolumetric contraction time. Ultrasound in Obstetrics Gynecology 2001;18(3):264–267.

[8] Shakespeare S, Crowe J, Hayes-Gill B, Bhogal K, James D. The information content of doppler ultrasound signals from the fetal heart. Medical and Biological Engineering and Computing 2001;39(6):619–626.

[9] Kupka T, Jezewski J, Matonia A, Horoba K, Wrobel J. Timing events in doppler ultrasound signal of fetal heart activity. In Engineering in Medicine and Biology Society, 2004. IEMBS'04. 26th Annual International Conference of the IEEE, volume 1. IEEE, 2004; 337–340.

[10] Khandoker AH, Kimura Y, Ito T, Sato N, Okamura K, Palaniswami M. Antepartum non-invasive evaluation of opening and closing timings of the cardiac valves in fetal cardiac cycle. Medical and Biological Engineering and Computing 2009;47(10):1075–1082.

[11] Marzbanrad F, Kimura Y, Funamoto K, Sugibayashi R, Endo M, Ito T, Palaniswami M, Khandoker AH. Automated estimation of fetal cardiac timing events from doppler ultrasound signal using hybrid models. Biomedical and Health Informatics IEEE Journal of 2014;18(4):1169–1177.

[12] Marzbanrad F, Kimura Y, Endo M, Oshio S, Funamoto K, Sato N, Palaniswami M, Khandoker A. Model based estimation of aortic and mitral valves opening and closing timings in developing human fetuses. Biomedical and Health Informatics IEEE Journal of 2014;PP. ISSN 2168-2194.

[13] Marzbanrad F, Khandoker A, Endo M, Kimura Y, Palaniswami M. A multi-dimensional hidden markov model approach to automated identification of fetal cardiac valve motion. In Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE. ISSN 1557-170X, Aug 2014; 1885–1888.

[14] Stroux L, Clifford G. The importance of biomedical signal quality classification for successful mhealth implementation. In 2014 Tech4Dev International Conference UNESCO Chair in Technologies for Development: What is Essential? June 2014; .

[15] Magenes G, Signorini M, Sassi R. Automatic diagnosis of fetal heart rate: comparison of different methodological approaches. In Engineering in Medicine and Biology Society, 2001. Proceedings of the 23rd annual international conference of the IEEE, volume 2. IEEE, 2001; 1604–1607.

[16] Sato M, Kimura Y, Chida S, Ito T, Katayama N, Okamura K, Nakao M. A novel extraction method of fetal electrocardiogram from the composite abdominal signal. Biomedical Engineering IEEE Transactions on 2007;54(1):49–58.

[17] Fleiss JL. Measuring nominal scale agreement among many raters. Psychological Bulletin 1971;76(5):378.

[18] Cardillo G. Fleiss's kappa: compute the Fleiss's kappa for multiple raters., 2007.

[19] Springer D, Brennan T, Zuhlke L, Abdelrahman H, Ntusi N, Clifford G, Mayosi B, Tarassenko L. Signal quality classification of mobile phone-recorded phonocardiogram signals. In Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on. IEEE, 2014; 1335–1339.

[20] Clifford G, Behar J, Li Q, Rezek I. Signal quality indices and data fusion for determining clinical acceptability of electrocardiograms. Physiological Measurement 2012; 33(9):1419.

[21] Kumar D, Carvalho P, Antunes M, Paiva R, Henriques J. Noise detection during heart sound recording using periodicity signatures. Physiological Measurement 2011; 32(5):599.

[22] Mitchell TM. Machine learning. 1997, volume 45. 1997.

[23] John GH, Langley P. Estimating continuous distributions in bayesian classifiers. In Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence. Morgan Kaufmann Publishers Inc., 1995; 338–345.

Address for correspondence:

Faezeh Marzbanrad
Department of Electrical and Electronic Engineering,
The University of Melbourne
Victoria 3010 Australia
f.marzbanrad@student.unimelb.edu.au