Estimation of Missing Data in Multi-channel Physiological Time-series by Reference Timing Channel and Average Substitution

P Langley, S King, K Wang, D Zheng, R Giovannini, M Bojarnejad, A Murray

Newcastle University & Freeman Hospital, Newcastle upon Tyne, UK

Physiological data continuously acquired from the clinical environment is often corrupted by noise, artifact or interruption so that sections of data are un-analysable. If sufficiently accurate estimations of these sections of data could be derived then a more complete analysis could be achieved. Our hypothesis was that reconstruction of missing data could be accurately achieved using average substitution if an appropriate reference timing channel without missing data was available.

Physiological data from the Computing in Cardiology Challenge 2010 were analysed. Reference timing was determined by an automatic beat detection algorithm from a channel without missing data. This was preferably the ECG channel, or if not available, a channel containing pulsatile data (eg blood pressure). For the channel with missing data the average beat across the recording was determined and this average provided the estimate of each missing beat at each beat time provided by the reference timing channel. Baseline offset estimate was achieved by ensuring the mean of the reconstructed data was equal to the mean of the 10 s of data prior to loss of signal. This algorithm would be ineffective for missing data which is non-pulsatile (eg respiration) so we implemented an algorithm based on PCA estimation of the respiratory data from the ECG channel. Results are shown in table 1.

	Event 1			Event 2			
Set	Basic	+Baseline	+Respiration	Basic	+Baseline	+Respiration	Attempted
A	50.1	52.9	54.3	65.1	65.1	66.5	100
B			67.8			78.0	100

Table 1. Score for each event and dataset for the basic algorithm, basic algorithm plus baseline estimation and the addition of the respiratory algorithm.

The final algorithm achieved scores of 54 (event1) and 67 (event 2) for set A and 68 (event 1) and 78 (event 2) for set B when providing estimates for missing data in all 100 recordings in each set. The algorithm does not take account of beat to beat changes in amplitude and consistently achieved higher scores based on correlation (event 2) than amplitude (event 1).