IOP PUBLISHING PHYSIOLOGICAL MEASUREMENT

Physiol. Meas. **33** (2012) 1491–1501 doi:10.1088/0967-3334/33/9/1491

Dynamic time warping and machine learning for signal quality assessment of pulsatile signals

Q Li[1,2] **and G D Clifford**[2]

[1] Institute of Biomedical Engineering, School of Medicine, Shandong University, Jinan, Shandong 250012, People's Republic of China
[2] Department of Engineering Science, Institute of Biomedical Engineering, University of Oxford, Oxford OX1 3PJ, UK

E-mail: gari@robots.ox.ac.uk

Received 21 February 2012, accepted for publication 26 June 2012
Published 17 August 2012
Online at stacks.iop.org/PM/33/1491

Abstract

In this work, we describe a beat-by-beat method for assessing the clinical utility of pulsatile waveforms, primarily recorded from cardiovascular blood volume or pressure changes, concentrating on the photoplethysmogram (PPG). Physiological blood flow is nonstationary, with pulses changing in height, width and morphology due to changes in heart rate, cardiac output, sensor type and hardware or software pre-processing requirements. Moreover, considerable inter-individual and sensor-location variability exists. Simple template matching methods are therefore inappropriate, and a patient-specific adaptive initialization is therefore required. We introduce dynamic time warping to stretch each beat to match a running template and combine it with several other features related to signal quality, including correlation and the percentage of the beat that appeared to be clipped. The features were then presented to a multi-layer perceptron neural network to learn the relationships between the parameters in the presence of good- and bad-quality pulses. An expert-labeled database of 1055 segments of PPG, each 6 s long, recorded from 104 separate critical care admissions during both normal and verified arrhythmic events, was used to train and test our algorithms. An accuracy of 97.5% on the training set and 95.2% on test set was found. The algorithm could be deployed as a stand-alone signal quality assessment algorithm for vetting the clinical utility of PPG traces or any similar quasi-periodic signal.

Keywords: artificial neural network, dynamic time warping, machine learning, multi-layer perceptron, photoplethysmograph, pulsatile signal, signal quality assessment

(Some figures may appear in colour only in the online journal)

1. Introduction

The photoplethysmograph (PPG) may not only be used as the source of arterial oxygen saturation (SaO_2) and heart rate (HR), but also as a simple and low-cost way of blood volume change detection in the microvascular bed of tissue, blood pressure and cardiac output estimation, respiration rate estimation and vascular assessment (Allen 2007). However, the PPG signal is easily disturbed by poor blood perfusion, ambient light and motion artifact (Hayes and Smith 1998, 2001). Such artifacts give rise to errors in the interpretation of the PPG signals in clinical physiological measurements, and can lead to numerous false alarms. In a recent study by Monstaerio *et al* (2012), apnea-related false desaturation alarm rates were shown to be as high as 85%.

Many signal processing methods have been used to suppress the artifacts, such as moving average filtering (Lee *et al* 2007), adaptive filtering (Graybeal and Petterson 2004, Chan and Zhang 2002, Relente and Sison 2002), wavelet transform (Sukanesh and Harikumar 2010, Addison and Watson 2010, Lee and Zhang 2003), independent component analysis (Kim and Yoo 2006, Yao and Warren 2005, Krishnan *et al* 2008a), high order statistics (Krishnan *et al* 2008b) and singular value decomposition (Reddy and Kumar 2007). However, the signal processing methodologies suffer from a lack of generality imposed by the implicit assumption that artifact corruption manifests itself as an additional signal component unrelated to the physiology either in the time, frequency or statistical domains (Hayes and Smith 2001). An alternative approach is to assess the signal quality of PPG waveform and consider analyzing only good-quality pulses. (Of course, the presence of poor quality waveforms can be considered useful information, such as a metric of physical activity, but the associated physiological information cannot be trusted.) Sukor *et al* (2011) used a waveform morphology analysis method to evaluate PPG signal quality when induced motion artifact occurred. By comparing with a manually annotated gold standard, the mean sensitivity, specificity and accuracy for beat detection were 89 $\pm$ 11%, 77 $\pm$ 19% and 83 $\pm$ 11%, respectively, on 104 fingertip PPG signals, acquired from 13 healthy people, conducted in a laboratory environment, containing varying degrees of purposely induced motion artifact. Gil *et al* (2010) and Monasterio *et al* (2012) used Hjorth parameters to assess PPG signal quality and Deshmane (2009) applied this to false electrocardiogram (ECG) arrhythmia alarms suppression in intensive care monitors. Although the Hjorth parameters provided an adequate method for identifying high-quality data segments, during arrhythmias the Hjorth parameters often identified PPG data associated with an arrhythmia as poor quality PPG. Moreover, the Hjorth parameters require a window much larger than a single beat, so temporal resolution is limited.

In this paper, we described a novel beat-by-beat PPG signal quality metric which uses a multi-layer perceptron (MLP) neural network to combine several individual signal quality metrics and physiological context to provide a probability of a pulse being acceptable for monitoring. One important component of our approach includes constructing an individual-specific template of an average beat. Dynamic time warping (DTW) (Keogh and Ratanamahatana 2005) was used to cope with the normal short-term nonstationary and nonlinear changes in height, width and overall morphology of each pulse due to changes in HR, cardiac output, manufacturer-specific hardware responses of sensors or software pre-processing requirements. (In the latter case, automatic changes in light intensity, amplifier gain or averaging may cause unusual distortions.) Furthermore, differences in individual recording modalities (such as the senor location or the method of attachment to the patient) and intra- and inter-individual variability in skin and cardiovascular state can lead to large differences in initial morphologies and dynamic changes. Simple template matching methods are therefore inappropriate, and an adaptive method of initializing on a given recording set-up, and tracking

the changes over time, is therefore required. For this reason, DTW has previously been employed in ECG segmentation and classification (Vullings *et al* 1998, Huang and Kinsner 2002). In this work, we use the DTW in a similar way to apply a nonlinear temporal stretching to fit the changing PPG beat with a dynamic beat template.

2. Methods

A database of 1055 expert-labeled beats drawn from 104 separate critical care recordings was used to develop the algorithm described in this work. For each recording, a template was first formed from the average of the 30 s of beats in the PPG waveform. The template was then updated by each new beat that is accepted (has an SQI above a given threshold). The degree of similarity between a given beat and a running template was then used as an index of signal quality.

However, since the DTW can fail in unexpected ways, it is not sufficient to just use this approach. A direct beat matching method without any preprocessing and also a matching based on linear resampling of the beat (to stretch or compress the beat to fit the length of the template) was also used. The correlation coefficient between the beat and the template was used as the signal quality index (SQI). Although the correlation coefficient can give a general match, it is insensitive to amplitudes, and indiscriminately accepts random square-wave noise. A clipping detection algorithm was therefore employed to detect the percentage of saturation to maximum or minimum value within each beat. These four measures of quality were then combined using a machine learning algorithm approach, which is described by Clifford *et al* (2011). Essentially, we learn the relationship between each of the signal quality measures by presenting the machine learning algorithm with hundreds of examples of high- and low-quality beats, and training the algorithm to classify the beats as high or low quality. This leads to a multivariate threshold set through rigorous experientially determined thresholds.

2.1. Beat detection

Beat detection was performed using wabp.c (an open source ABP beat detector (Zong *et al* 2003) from www.physionet.org) with a time and amplitude threshold adjustment to fit PPG beat width and height. Specifically, we changed the slope width of the rising edge of the beat from 130 to 170 ms and extended the eye-closing period after each detected beat from 250 to 340 ms to avoid double detection of the possible secondary peak of a PPG beat. The length of a PPG beat was delimited by the fiducial marks at the onset of the current beat and the onset of the next beat. If no beat was found 3 s after the onset of any given beat, then the end of the beat window was truncated to 3 s.

2.2. Initial template generation

A PPG beat template was initially generated by averaging every beat in a window of 30 s. The PPG signals are assumed to be quasi-periodic, and so autocorrelation of each 30 s of data was taken and the length (L) between two main peaks of the autocorrelation sequence was used to determine the average period of PPG beats. The length of the PPG template was then set to be L. To derive the first template (T_1), we averaged all the beats in the 30 s window with each beat beginning at the fiducial mark (onset of the beat) and ending at the length of the template. The correlation coefficients (C) between T_1 and each beat in the 30 s window were then calculated (Clifford 2002). Any beat with $C < 0.8$ was removed from the template, and the average beat was recalculated from the remaining beats to generate the second template (T_2). If more than

half of the beats were removed by the process, T_2 was deemed untrustworthy, and the template from the previous window was used instead. If no previous window is available, the next 30 s were used. Template updating can then be performed on a beat-by-beat basis, but only after classification of a new incoming beat is performed, which requires several other beat analysis metrics first as described below.

2.3. DTW of the PPG beat

As described earlier, a nonlinear time-base stretching of each beat is sometimes required before correlating with the beat template, in order to allow for nonlinear and nonstationary changes in the beat morphology. This was achieved through DTW. Suppose we have two time series, T and B, of lengths n and m, respectively, where

$$\mathrm{T} = t_1, t_2, \ldots, t_i, \ldots, t_n \tag{1}$$

$$\mathrm{B} = b_1, b_2, \ldots, b_j, \ldots, b_m. \tag{2}$$

To align two sequences using DTW, an $n \times m$ distance matrix (D) is constructed where the (ith, jth) element of the matrix contains the distance $d(t_i, b_j)$ between the two points t_i and b_j. Each matrix element (i, j) corresponds to the alignment between the points t_i and b_j. The aim of DTW is to find an optimal path from (0, 0) to (n, m) and minimize the cumulative distance of the path.

Defining T as the template of PPG and B as a PPG beat, we first transform the template and the beat to short line sequences using a piecewise linear approximation (PLA) algorithm (Koski 1996). The distance between each short line pair ($d(t_i, b_j)$) is then defined as the absolute difference between the slopes of each short line. A cumulative distance up to lines i and j, $c_{i,j}$, is then defined by

$$c_{i,j} = \min \begin{cases} c_{i-1,j} + d(t_i, b_j) l(t_i) \\ c_{i-1,j-1} + d(t_i, b_j)(l(t_i) + l(b_j)) \\ c_{i,j-1} + d(t_i, b_j) l(b_j), \end{cases} \tag{3}$$

where $l(t_i)$ and $l(b_j)$ are the durations of lines t_i and b_j in the time series, respectively. The optimal path can be achieved by selecting the path with the minimum cumulative distance. Figure 1 shows an example of the PPG template and beat sequences, optimal warping path and the resulting alignment.

2.4. Signal quality metrics for PPG

Four individual SQIs were initially defined as follows.

2.4.1. Direct matching SQI. We selected the sampling point series of each beat within the 30 s window, beginning at the fiducial mark and ending at the length of the template (L). Then calculated the correlation coefficient with the template as the direct matching SQI (SQI_1). We set any negative value of correlation coefficient (negative correlation) to zero, so the value of SQI ranges between 0 and 1 inclusively.

2.4.2. Linear resampling SQI. We selected each beat between two fiducial marks and linearly stretched (if the length of the beat is shorter than L) or compressed (if it is longer) the beat to the length of the template. Then calculated the correlation coefficient as the linear resampling SQI (SQI_2). Again, the SQI value was rounded to a non-negative number.

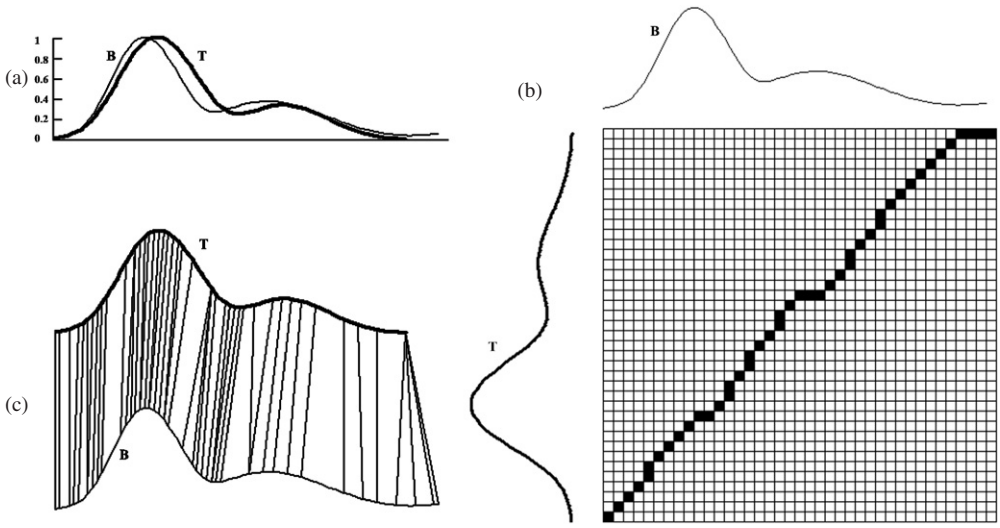

Figure 1. An example of the DTW procedure. (a) The PPG beat template (T—bold line) and a PPG beat (B—soft line). (b) To align T and B, a warping matrix was constructed and the optimal warping path was shown with solid squares. (c) The resulting alignment flow.

2.4.3. Dynamic time warping SQI. Using DTW, we resample the beat to the length L and calculate the correlation coefficient as the dynamic time warping SQI (SQI_3). Non-negative rounding is again applied.

2.4.4. Clipping detection SQI. Periods of saturation to a maximum or a minimum value were determined within each beat. A hysteresis threshold (of 1 normalized unit) to determine the smallest fluctuation that should be ignored was defined. Such samples are defined to be 'clipped'. The percentage of the beat that is *not* clipped is defined to be the clipping detection SQI (SQI_4).

2.5. Data sources

As there is no annotated PPG database published, we trained and evaluated our algorithm using an annotated PPG dataset developed by the PhysioNet team (Goldberger *et al* 2000) taken from the MIMIC II database (Saeed *et al* 2002). The dataset includes 1437 signal quality annotations of each channel including ECG, arterial blood pressure (ABP) and PPG from 104 independent adult critical care stays. Two independent annotators graded the signal quality based on the waveform around the time when arrhythmia alarm of monitors occurs. Disagreement was adjudicated by a third expert. There are two types of arrhythmia alarms in the dataset: asystole and ventricular tachycardia (VT). The types of annotations for signal quality were as follows: good (1), bad (0) and uncertain (other). We selected only the annotations with a value of 1 (good) or 0 (bad) to be used in this study. The distribution of these annotations for the dataset is shown in table 1.

Data were then split into separate training and testing groups. Patients in the dataset were sorted in ascending order of the number of annotations they possessed and every odd numbered patient (in the sorted list) was placed in the training and every even numbered patient in the test set. Each set therefore had an equal number of patients (52) and an approximately equal number of annotations, as shown in table 2.

Table 1. Summary of the expert annotations in the dataset.

		PPG annotations				
Alarm type	Patients	Good	Bad	Uncertain	Total	Used (good + bad)
Asystole	54	177	75	97	349	252
VT	88	648	155	285	1088	803
Total	104	825	230	382	1437	1055

Table 2. Summary of the annotations in training and test datasets.

Dataset	Good quality	Bad quality	Total
Training	427	127	554
Test	398	103	501
Total	825	230	1055

2.6. Data fusion approaches

Two methods for fusing the signal quality information were compared: one based on simple logic and one using an optimized multivariate classifier (the MLP).

2.6.1. Simple heuristic fusion of the SQI matrices. The four signal quality indices were fused into one (qSQI) and used to classify each beat in the dataset. The fusion equation was constructed in an ad hoc manner as follows:

$$q\,\mathrm{SQI} = \begin{cases} \text{Excellent (E)} & \text{if All of the 4 } \mathrm{SQI}_i \geqslant 0.9 \\ \text{Acceptable(A)} & \begin{cases} \text{if 3 of the 4 } \mathrm{SQI}_i \geqslant 0.9 \text{ OR} \\ \text{if All of the 4 } \mathrm{SQI}_i \geqslant 0.7 \text{ OR} \\ \text{if median}(\mathrm{SQI}_1, \mathrm{SQI}_2, \mathrm{SQI}_3) \geqslant 0.8 \text{ and } \mathrm{SQI}_1 \geqslant 0.5 \text{ and } \mathrm{SQI}_4 \geqslant 0.7 \end{cases} \\ \text{Unacceptable } (U) & \text{Otherwise,} \end{cases} \tag{4}$$

where the coefficients 0.9, 0.8, 0.7 and 0.5 are arbitrary and set empirically through trial and error. Although these coefficients could be optimized, it is unlikely that the logic is optimal, and so an exhaustive search of possible logical combinations and thresholds was not performed. Rather, qSQI was defined to provide a baseline for a more principled approach. To convert the categorical outputs to numerical outputs, we mapped E or A to a value of unity, and U to a value of zero.

To evaluate the performance of the algorithm, we chose an analysis window of 6 s, beginning at 5 s before the asystole or VT alarm onset. (This was approximately the segment of data which was used to make the SQI annotation by the experts.) An extra window of 30 s before the alarm fiducial mark was used to generate the 'normal' beat template. The mean qSQI ($q\mathrm{SQI}_{\mathrm{mean}}$) of all the beats within the analysis window was calculated. At the training stage, we selected a good-quality threshold ($q\mathrm{SQI}_{\mathrm{th}}$) to achieve the best classification accurate rate for the training set. If $q\mathrm{SQI}_{\mathrm{mean}} \geqslant q\mathrm{SQI}_{\mathrm{th}}$, we set the SQI to 1; otherwise we set the SQI to 0 in order to compare with the gold standard expert annotations and calculate the accuracy. To select the best $q\mathrm{SQI}_{\mathrm{th}}$, we varied its value between 0 and 1 in steps of 0.01 and calculated the classification accuracy at each point. The best $q\mathrm{SQI}_{\mathrm{th}}$, which resulted in the highest accuracy, was then used to classify the test set.

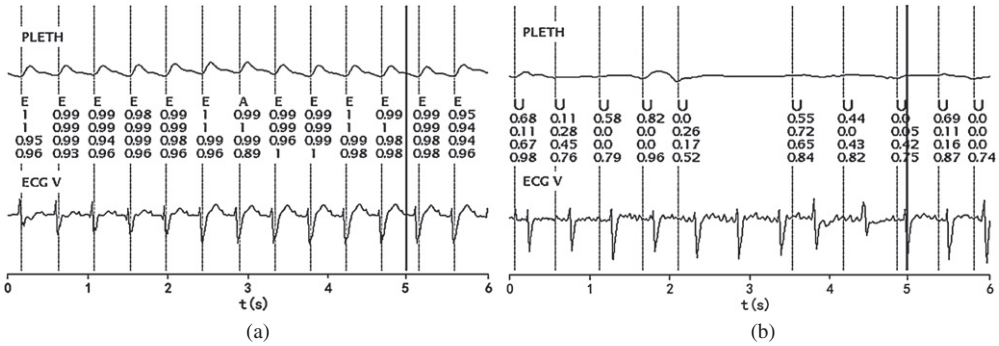

Figure 2. An example of SQI matrices and simple fusion of PPG from the evaluation dataset. (a) Annotated as E or A (good quality), (b) annotated as U (bad quality). Each plot shows two channels of signal: PPG (PLETH) and ECG (ECG V). The ECG is provided for visual reference only and is not used. Each detected PPG beat was marked by a dotted line and accompanied by a column of five annotations corresponding to the individual beat's values of qSQI, (categorical: E, A or U), and the numerical values of SQI_1, SQI_2, SQI_3 and SQI_4, respectively. Note that equation (4) was applied to SQI_1 through SQI_4 to determine qSQI.

2.6.2. Machine learning for quality estimation. We selected two groups of input variables to present to the MLP. The first group included the four SQI metrics (SQI_1, SQI_2, SQI_3 and SQI_4). For each SQI metric, we calculated the mean SQI of the beats within the 6 s analysis window. The second group used six variables, including the four SQI matrices, the simple fusion (qSQI) and the number of beats detected within the window (N_{beats}). The rationale for adding the number of beats as an input was that we expect the noise and abnormality of the signal to manifest differently at different HRs. The rationale for including qSQI as a feature is that, if it proves to be a useful approach, then the highly nonlinear structure of the metric's logic would be difficult to reproduce without much larger numbers of training patterns.

Therefore, the architecture of the MLP was 4-N-1 or 6-N-1, where the number of hidden nodes, N, had to be optimized, and the input was fixed to the number of features as described above. The output was simply a single node providing an estimate of the class (1 or 0). A sigmoid activation function was used on the hidden layer and the MLP neural network training used the Levenburg–Marquardt algorithm (Moré 1978). The stopping criteria were as follows: a maximum of 200 epochs, an error $\leqslant 10^{-5}$ or a gradient $\leqslant 10^{-5}$. Since the MLP requires an independent validation set to prevent over-training, the training set was further divided into subsets 70% for training, 25% for validation and 5% for pre-testing at random. The validation set was used to test the optimal number of nodes in the hidden layer. This was chosen to be the number which provided the highest accuracy within the range of $N = 2$–20. (Using more than 20 hidden nodes would likely lead to extreme over-fitting for our given dataset.)

3. Results

3.1. SQI metrics of PPG

The four SQI metrics quantify different characteristics and the simple fusion of the SQI matrices (qSQI) classifies the signal quality of each PPG beat into three levels: extremely high quality (E), moderate quality (A) and untrustworthy (U). Figure 2 shows two parts of PPG from the evaluation dataset with four SQI metrics and the simple fusion classification. Each PPG beat onset is marked by a dotted line and the alarm onset is marked by a solid line at the 5th second.

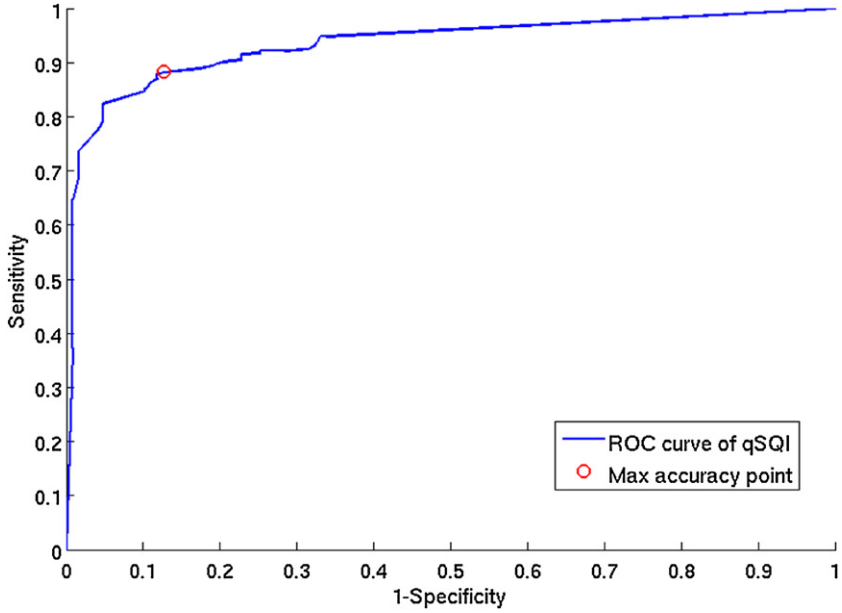

Figure 3. ROC curve of the qSQI algorithm derived by varying $q\text{SQI}_{\text{th}}$ across the training set. The circle indicates the position of maximum accuracy (88.1% in the training set).

Table 3. Performances of heuristic and ML approaches. The best performing algorithm (on the independent test set) is indicated in bold.

		Training performance (%)				Test performance (%)				
Method	No of Inputs	Acc	Se	SP	PPV	Acc	Se	SP	PPV	Notes
qSQI	1	88.1	88.3	87.4	95.9	91.8	94.7	80.6	95.0	$q\text{SQI}_{\text{th}} = 0.36$
MLP	6	97.5	98.4	94.5	98.4	**95.2**	**99.0**	**80.6**	**95.2**	Hidden nodes: 10
MLP	4	97.1	98.6	92.1	97.7	92.4	96.7	75.7	93.9	Hidden nodes: 10

3.2. Evaluation results

3.2.1. Result of qSQI. Using the training set, we varied the value of $q\text{SQI}_{\text{mean}}$ above which data were considered to be good quality and calculated the receiver operating characteristic (ROC) curve (figure 3). The $q\text{SQI}_{\text{th}}$ which gave the best classification accuracy was $q\text{SQI}_{\text{th}} = 0.36$, which resulted in an accuracy of 88.1% (488 correctly classified out of 554) on the training set. Using this threshold, the accuracy on the test set was found to be 91.8% (460 correctly classified out of 501).

3.2.2. Results of machine learning for classifying quality. In contrast to thresholding on qSQI, the machine learning algorithm approach provides a multivariate threshold. Figure 4 shows the ROC curves of the MLP algorithm. The MLP neural network with six inputs gives the best performance with an accuracy of 97.5% (540 of 554) on the training set and 95.2% (477 of 501) on the test set.

The full performances of different quality estimation methods are shown in table 3.

Finally, in order to test the multivariate marginal information increase of each input variable, we retrained the MLP algorithm for all combinations of five of the six input variables. Table 4 shows the performance of each of these combinations. The highest accuracy on test

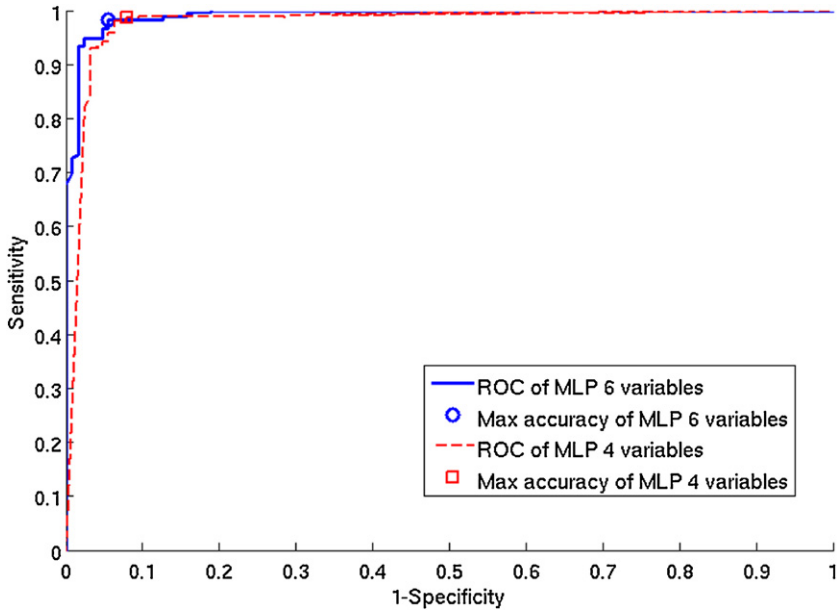

Figure 4. ROC curves of the MLP algorithms for the training set with operating points of maximal accuracy indicated.

Table 4. Performances of any possible five inputs of the MLP algorithm. The best performing algorithm (on the independent test set) is indicated in bold.

	Training performance (%)				Test performance (%)				No of hidden
Inputs	Acc	Se	SP	PPV	Acc	Se	SP	PPV	nodes
qSQI, SQI_1, SQI_2, SQI_3, SQI_4	97.3	99.3	90.1	97.3	91.2	98.0	65.1	91.6	13
qSQI, SQI_1, SQI_2, SQI_3, N_{beats}	97.7	98.8	93.7	98.1	**94.6**	**97.0**	**85.4**	**96.3**	14
qSQI, SQI_1, SQI_2, SQI_4, N_{beats}	97.1	99.3	89.8	97.0	94.6	98.0	81.6	95.4	6
qSQI, SQI_1, SQI_3, SQI_4, N_{beats}	98.4	99.1	96.1	98.8	93.6	97.2	79.6	94.9	19
qSQI, SQI_2, SQI_3, SQI_4, N_{beats}	98.7	99.8	95.3	98.6	92.0	96.5	74.8	93.7	19
SQI_1, SQI_2, SQI_3, SQI_4, N_{beats}	98.6	98.6	98.4	99.5	94.0	96.7	83.5	95.8	18

data was 94.6% with variables qSQI, SQI_1, SQI_2, SQI_3 and N_{beats}, which is marginally lower than the best performance of 95.2%, with a small drop in sensitivity (Se), from 99% to 97%, but a large increase in specificity (SP) and a marginal increase in positive predictivity (PPV). We note that the number of hidden nodes found for this performance is relatively high (14). A similar performance was found using only six hidden nodes qSQI, SQI_1, SQI_2, SQI_4 and N_{beats}, indicating that much complementary information exists between each metric.

4. Discussion

The multivariate 'voting' threshold provided by the machine learning approach is clearly superior to the single parameter thresholding on the SQI metrics, although only if a good choice of ML algorithm is made. Although other ML algorithms could be used, the flexibility of the neural network and its simple on-line implementation make it a good choice if large numbers of training patterns are available (and in fact, in tests not published here, a support vector machine produced marginally worse results). Of the tested approaches, the MLP using

all six quality measures provided the best performance, with 95% accuracy on an independent (unseen) test set. Although this is an impressive accuracy, and similar to recent results on the ECG quality analysis we performed with a paradigmatically similar approach (Clifford *et al* 2011), it must be noted that the weights of our trained MLP are specific to the type of data on which it was trained. In other words, to extend this system to other data and rhythms (outside of asystole and ventricular tachycardia), the MLP must be retrained. This, of course, is not an issue as long as accurately labeled data are available. It should also be noted that there is some ambiguity in interpreting the 95% accuracy of our system in as much as it is not known what level of accuracy would be needed in a particular circumstance or application. For example, such an accuracy may be entirely sufficient to detect HRs (and reduce false alarms such as bradycardia, asystole and tachycardia), but may not be sufficient to determine if we could trust an apnea alarm resulting from an analysis of a respiratory trace derived from the PPG, or a desaturation alarm. In subsequent studies, we will attempt to assess such questions.

By systematically removing each of the six input features, we see that the accuracy always drops, by between 0.6% and 3.8% from the six-input performance of 95%. This shows that every quality metric provides some improvement in a multivariate sense with N_{beats} providing the most additional marginal information and SQI_4 providing the least. This is as we would expect, since N_{beats} (which is proportional to HR) is the most independent input parameter and a measurement of saturation (SQI_4) may be redundant compared to the template matching. Moreover, the interpretation of each of the SQIs should be HR dependent.

A final note concerns the choice of features in this study, which were based on intuition and experience. However, the features are not exhaustive and a much wider variety of features could be tested as described in this work, or by adding in a feature selection approach such as a genetic algorithm.

5. Conclusion

We have described an effective system (with 95% accuracy on unseen test data) which could be deployed as a stand-alone signal quality assessment algorithm for vetting the clinical utility of PPG signals. Applications range from false alarm suppression to improving estimates of derived physiological parameters such as HR, respiration, oxygen saturation, pulse transit time and peripheral circulatory changes. Moreover, the algorithm presented here is quite general and could be retrained and applied to any periodic or quasi-periodic signal such as continuous blood pressure.

Acknowledgments

The authors gratefully acknowledge funding for this research from Mindray North America. The authors would also like to thank the Laboratory for Computational Physiology at MIT for providing the annotated data for this study.

References

Addison P S and Watson J N 2010 Signal processing techniques for determining signal quality using a wavelet transform ratio surface *US Patent* No. 2010/0298728 A1

Allen J 2007 Photoplethysmography and its application in clinical physiological measurement *Physiol. Meas.* **28** R1–39

Chan K W and Zhang Y T 2002 Adaptive reduction of motion artifact from photoplethysmographic recordings using a variable step-size LMS filter *Proc. IEEE Sensors* vol 2 pp 1343–6

Clifford G D 2002 Signal processing methods for heart rate variability *DPhil Thesis* Oxford University, Oxford, UK

Clifford G D, Lopez D, Li Q and Rezek I 2011 Signal quality indices and data fusion for determining acceptability of electrocardiograms collected in noisy ambulatory environments *Comput. Cardiol.* **38** 285–8

Deshmane A V 2009 False arrhythmia alarm suppression using ECG, ABP, and photoplethysmogram *MS Thesis* MIT, Cambridge, MA, USA

Gil E, Bailon R, Vergara J and Laguna P 2010 PTT variability for discrimination of sleep apnea related decreases in the amplitude fluctuations of PPG signal in children *IEEE Trans. Biomed. Eng.* **57** 1079–88

Goldberger A L, Amaral L A N, Glass L, Hausdorff J M, Ivanov P C, Mark R G, Mietus J E, Moody G B, Peng C K and Stanley H E 2000 PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals *Circulation* **101** e215–20

Graybeal J M and Petterson M T 2004 Adaptive filtering and alternative calculations revolutionizes pulse oximetry sensitivity and specificity during motion and low perfusion *Proc. 26th Annu. Int. Conf. IEEE EMBS* pp 5363–6

Hayes M J and Smith P R 1998 Artifact reduction in photoplethysmography *Appl. Opt.* **37** 7437–46

Hayes M J and Smith P R 2001 A new method for pulse oximetry possessing inherent insensitivity to artifact *IEEE Trans. Biomed. Eng.* **48** 452–61

Huang B and Kinsner W 2002 ECG frame classification using dynamic time warping *Proc. 2002 IEEE Canadian Conf. on Electrical and Computer Engineering* pp 1105–10

Keogh E and Ratanamahatana C A 2005 Exact indexing of dynamic time warping *Knowledge and Information Systems* vol 7 (London: Springer) pp 358–86

Kim B S and Yoo S K 2006 Motion artifact reduction in photoplethysmography using independent component analysis *IEEE Trans. Biomed. Eng.* **53** 566–8

Koski A 1996 Segmentation of digital signals based on estimated compression ratio *IEEE Trans. Biomed. Eng.* **43** 928–38

Krishnan R, Natarajan B and Warren S 2008a Motion artifact reduction in photopleythysmography using magnitude-based frequency domain independent component analysis *Proc. 17th Int. Conf. on Computer Communications and Networks ICCCN '08* pp 1–5

Krishnan R, Natarajan B and Warren S 2008b Analysis and detection of motion artifact in photoplethysmographic data using higher order statistics *IEEE Int. Conf. on Acoustics, Speech and Signal Processing* pp 613–6

Lee C M and Zhang Y T 2003 Reduction of motion artifacts from photoplethysmographic recordings using a wavelet denoising approach *IEEE EMBS Asian-Pacific Conf. on Biomed. Eng.* pp 194–5

Lee H W, Lee J W, Jung W G and Lee G K 2007 The periodic moving average filter for removing motion artifacts from PPG signals *Int. J. Control Autom. Syst.* **5** 701–6

Monasterio V, Burgess F and Clifford G D 2012 Robust classification of neonatal apnoea-related desaturations *Physiol. Meas.* **33** 1503–16

Moré J J 1978 The Levenberg–Marquardt algorithm: implementation and theory *Numerical Analysis (Lecture Notes in Mathematics* vol 630*)* ed G A Watson (Berlin: Springer) pp 105–16

Reddy K A and Kumar V J 2007 Motion artifact reduction in photoplethysmographic signals using singular value decomposition *Proc. IEEE Instrumentation and Measurement Technology Conf. IMTC* pp 1–4

Relente A R and Sison L G 2002 Characterization and adaptive filtering of motion artifacts in pulse oximetry using accelerometers *Proc. 2nd Joint EMBS/BMES Conf.* vol 2, pp 1769–70

Saeed M, Lieu C, Raber G and Mark R G 2002 MIMIC II: a massive temporal ICU patient database to support research in intelligent patient monitoring *Comput. Cardiol.* **29** 641–4

Sukanesh R and Harikumar R 2010 Analysis of photo-plethysmography (PPG) signals with motion artifacts (Gaussian noise) using wavelet transforms *Biomed. Softw. Comput. Hum. Sci.* **16** 135–9

Sukor J A, Redmond S J and Lovell N H 2011 Signal quality measures for pulse oximetry through waveform morphology analysis *Physiol. Meas.* **32** 369–84

Vullings H J L M, Verhaegen M H G and Verbruggen H B 1998 Automated ECG segmentation with dynamic time warping *Proc. 20th Annu. Conf. IEEE EMBS* vol 20, pp 163–6

Yao J and Warren S 2005 A short study to assess the potential of independent component analysis for motion artifact separation in wearable pulse oximeter signals *IEEE-EMBS 27th Annu. Int. Conf. of the Engineering in Medicine and Biology Society* pp 3585–8

Zong W, Heldt T, Moody G B and Mark R G 2003 An open-source algorithm to detect onset of arterial blood pressure pulses *Comput. Cardiol.* **30** 259–62