

Photoplethysmograph Quality Estimation Through Multichannel Filtering

Ikaro Silva, *Member, IEEE*, Joon Lee, *Member, IEEE*, and Roger Mark, *Fellow, IEEE*

Abstract—Information about the quality of a recorded physiological waveform can be valuable for the detection of critical medical conditions. This work presents a new point-by-point signal quality index (SQI) based on adaptive multichannel prediction which does not rely on ad-hoc morphological feature extraction from the target waveform. An application of the SQI to photoplethysmograph waveforms showed that the SQI is monotonically related to SNR (simulated by adding white noise) and subjective human quality assessment of 1,313 waveform epochs. A receiver-operating-characteristic (ROC) curve analysis, with the human “bad” quality label as negative and the human “good” quality label as positive, yielded an area under the ROC curve of 0.863. For photoplethysmograph waveforms, a SQI greater than 0.8 seems in general to be indicative of good signal quality.

I. INTRODUCTION

Estimation of the quality of a physiological signal has several important applications; the estimated quality can be used to decrease the number of false detections of critical conditions, detect sensor displacement, or process and extract clinically relevant features. While several ad-hoc algorithms exist for estimating quality of specific types of physiological signals, little work has been done in developing a generic signal quality index for physiological signals based on multichannel recordings. Qualitative observations of typical clinical multichannel recordings, that include electrocardiogram, photoplethysmograph, arterial blood pressure, and respiration, indicate a general coupling between these different physiological measurements. This work proposes a new Signal Quality Index (SQI) for photoplethysmographic data based on coupling information estimated from concurrent waveforms.

The multichannel adaptive filter (MCAF) [1] originally developed for the 2010 PhysioNet Challenge [2] was used to estimate the degree of coupling between the desired channel and other concurrent waveforms. This filter utilizes a combination of linear adaptive filters to predict the desired signal using the other concurrent channels. Using the MCAF, a SQI was derived based on the explicit assumption that “good” quality signals will yield a low MCAF tracking error, and that “bad” quality signals will yield a high MCAF prediction error.

I. Silva is a Postdoctoral Associate in the Laboratory for Computational Physiology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA ikaro@mit.edu

J. Lee is a Postdoctoral Associate in the Laboratory for Computational Physiology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA joonlee@mit.edu

R. Mark is a Distinguished Professor of Health Sciences and Technology and Professor of Electrical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA rgmark@mit.edu

II. METHODS

A. Dataset and Human Annotated Signal Quality

MIMIC II [3], a public database consisting of thousands of intensive care unit waveforms, was used to extract 1,313 epochs for this study. Each recorded epoch consisted of an $N \times M$ measurement matrix

$$X = [\mathbf{x}_1 \ \mathbf{x}_2 \ \dots \ \mathbf{x}_M], \quad (1)$$

where each column, \mathbf{x}_i , represents a $N \times 1$ vector (a single channel). The channel to be predicted was defined as the target channel \mathbf{x}_T .

The numbers of channels available for all of the 1,313 epochs varied between 6 and 12. All epochs included the following physiological signals: respiration (RESP) and photoplethysmograph (PLETH). Additional channels included a combination of common recordings such as: arterial blood pressure (ABP), central venous pressure (CVP), pulmonary arterial pressure (PAP), right atrial pressure (RAP), and the following electrocardiogram (ECG) leads: I, II, III, V, AVR, AVL, AVF, and MCL. Each epoch was 10 minutes long, sampled at 125 Hz at 10-bit resolution. The epoch’s end time (i.e., the region to be labeled) corresponded to an arrhythmia alarm triggered by an intensive care bedside monitor. This region was chosen due to the clinical significance of these events and the potential for signal quality estimates to reduce false alarm rates. Moreover, most of these alarms were triggered due to changes in the measurements, thus the selected epochs are likely to include non-stationary statistics. Selection of our epochs based on these alarm regions ensures that the SQI algorithm is being tested under realistic and relevant clinical conditions.

Two human annotators classified the signal quality of the PLETH channel into three categories: “good”, “maybe” and “bad”. A third human annotator adjudicated disagreement between the two initial annotators. The third annotator was the most senior and experienced, and his label overrode any disagreement. The labeling of “maybe” meant that the peaks of the signal were clearly visible, and that further processing such as filtering, could help clean the signal. Although each epoch was labeled at the end, the annotators were still allowed to analyze additional data beyond the end of the selected 10-minute region. In general, the annotators made their decision on the quality of the signal based on a few seconds before and after the 10 minute mark. Table I shows the general statistics of the dataset.

TABLE I
DATASET QUALITY DESCRIPTION.

	Good	Maybe	Bad	Total
Number of Epochs	823	260	230	1313

B. Adaptive Filtering And Prediction

The MCAF prediction algorithm consisted of a bank of M Gradient Adaptive Laguerre Lattice (GALL) filters [4] followed by a Kalman filter [5] that combined the individual responses to generate a final estimate (Figure 1). The Kalman filter covariance matrix was set to the identity matrix (assuming equal “state” noise levels on the output of the individual GALL estimates). The first 9 minutes and 30 seconds of the epoch were used to optimize the forgetting factor for all the filters (including the Kalman filter), as well as the single pole location of the GALL filters. The mean square difference between the target signal and the Kalman estimate was used as a cost function. The channel to be tracked was defined as the target channel (which was the PLETH channel in this particular case). The input to the T th GALL filter was a 30-second delayed version of the target signal [1].

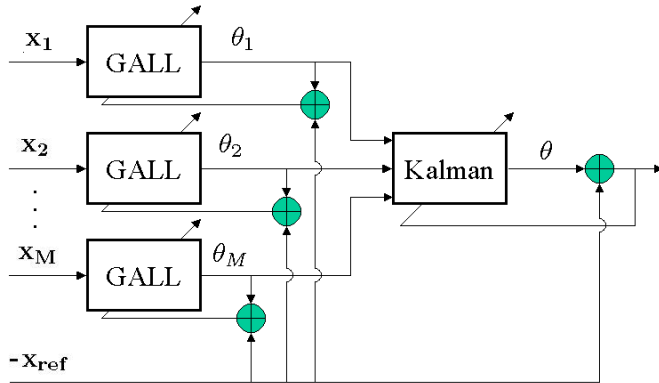


Fig. 1. MCAF Prediction Overview

C. Estimating Signal Quality

The signal quality index for a single channel was estimated in two steps. In the first step a preliminary point-by-point SQI, $SQI_p[n]$, was derived from the predicted MCAF signal, θ , as

$$SQI_p[n] = \frac{1}{1 + \frac{(\theta[n] - x_T[n])^2}{\theta[n]^2}}. \quad (2)$$

Note that $SQI_p[n]$ is bounded between 0 and 1 (with perfect estimation of a signal yielding an $SQI_p[n]$ of 1). The second and final step consisted of gating $SQI_p[n]$ with a masking function defined as 0 for regions where the first derivative of x_T was 0 (otherwise the masking function was defined as 1). The product of $SQI_p[n]$ with this masking function was then low-pass filtered using a five second moving average filter

$$SQI[n] = \frac{1}{L} \sum_{i=n-L}^n SQI_p[i] \cdot gate[i], \quad (3)$$

where

$$gate[n] = \begin{cases} 0 & \text{if } x_T[n] = x_T[n-1] \\ 1 & \text{otherwise} \end{cases} \quad (4)$$

Note that the gating function, $gate[n]$, is 0 in regions where the first derivative of $x_T[n]$ is 0, which is likely to occur with clipping artifacts, no signal, or a constant DC output (all these cases were defined as low quality).

D. Gaussian Noise Simulation

In order to characterize the SQI as a function of signal-to-noise ratio (SNR) for noise source uncorrelated across channels, a series of simulations was generated by adding stationary white Gaussian noise to a PLETH signal labeled by the annotators as “good” (Figure 2). This particular epoch had the following channels: CVP, RESP, ECG AVR, ECG II, ABP, ECG III, and ECG V in addition to the target channel (for ease of visualization, the ECG leads III and V were omitted from Figure 2). White Gaussian noise was added to the target signal at the following SNR levels (dB) : -30, -20, -10, -5, 0, 5, 10, 20, and 30. Figure 2 shows the case of a SNR of -10 dB. Note that noise was added to the target signal only, and the original PLETH signal is displayed for comparison (it was never used by the MCAF algorithm).

E. SQI comparison with Human Labels

A comparison of the estimated SQI and human labeling was performed using two different methods. In the first method, statistics of the estimated SQI (mean and standard error) were conditioned on the three label categories and analyzed. In the second method, a receiver operating characteristic (ROC) curve and its area were computed. The ROC was generated using the estimated SQI as the dependent variable, the “good” label as positive, and the “bad” label as negative. Due to the binary nature of the ROC analysis, only the two extreme labels were chosen and the “maybe” class was ignored.

III. RESULTS

An example of the SNR simulation in the MCAF prediction algorithm at -10 dB SNR is shown in Figure 2. The estimated SQI is a monotonic function of SNR, yielding values of less than 0.9 for SNRs less than 0. Figure 3 shows the results comparing estimated SQI (Equation (3)) and human labeling of the epochs. These results are also described in Table II. The area under the ROC of Figure 3, was 0.86, showing a reasonable separation between the two classes. Figure 4 shows the results of the SQI estimation under the various simulated SNR conditions.

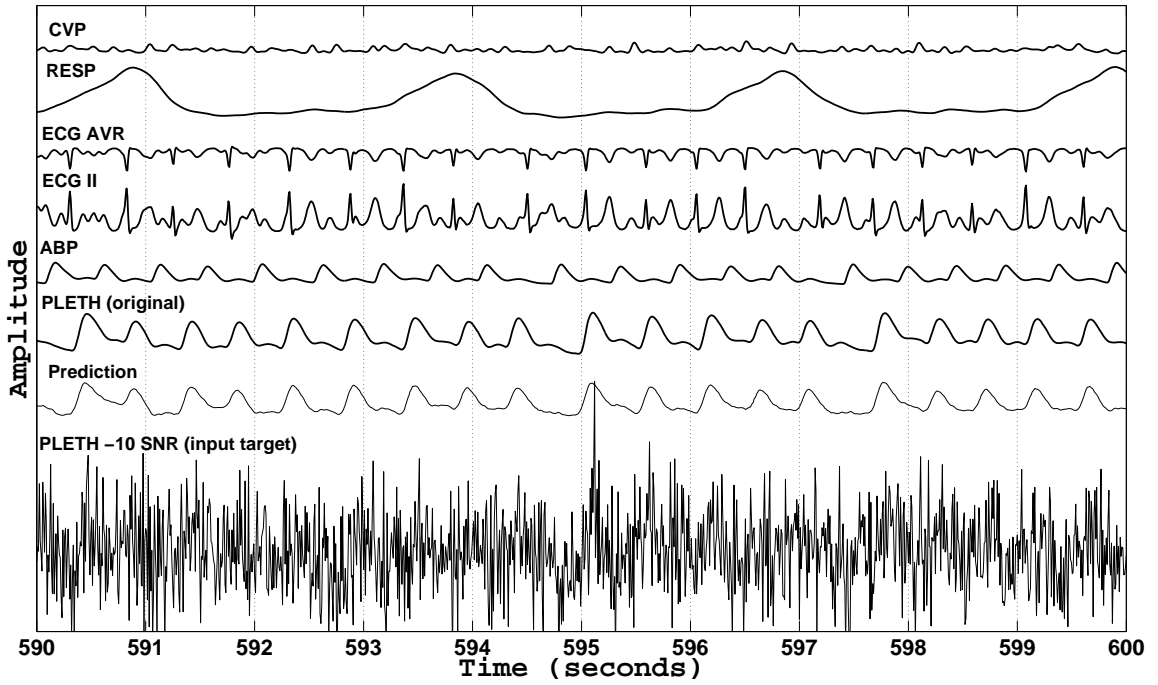


Fig. 2. Example of the MCAF's output (Prediction) with a SNR of -10 dB input target (PLETH -10 SNR). The input target ((PLETH -10 SNR) was used as both x_T and x_{ref} on the MCAF algorithm (Fig. 1). The original signal (PLETH original) is displayed for comparison purposes only and was never used. Some of the concurrent channels are shown on the top.

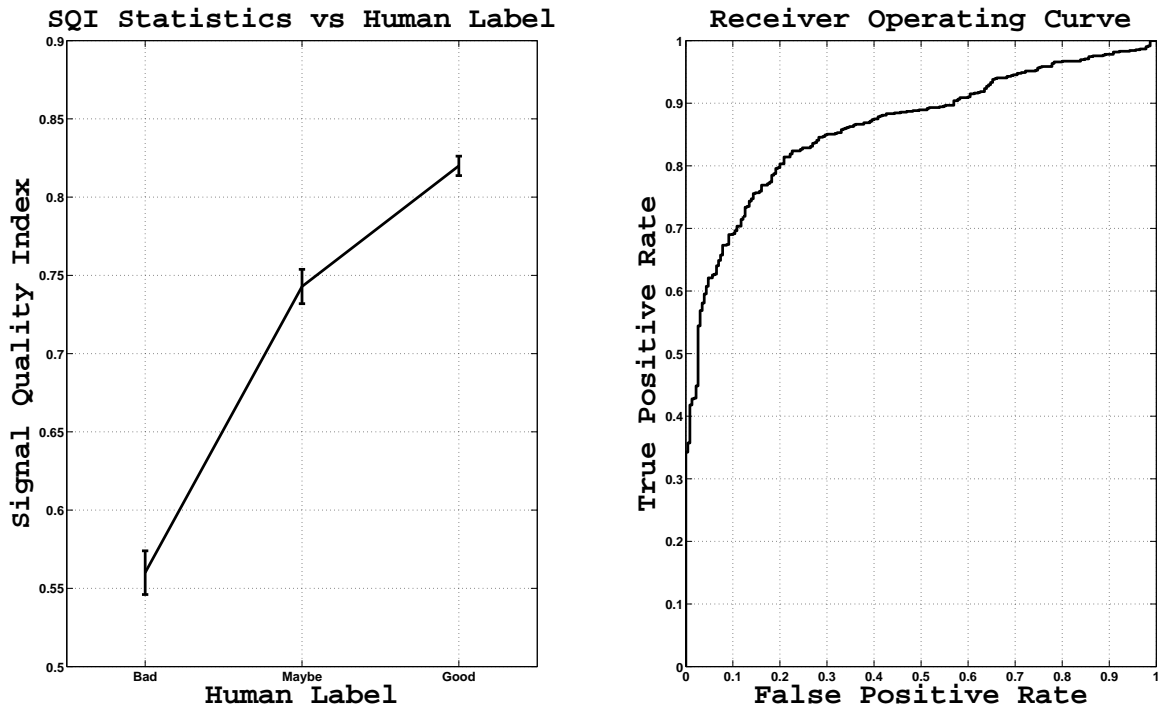


Fig. 3. Estimated SQI statistics conditioned on human labels (error bars represent plus or minus one standard error). The area under the ROC was 0.86.

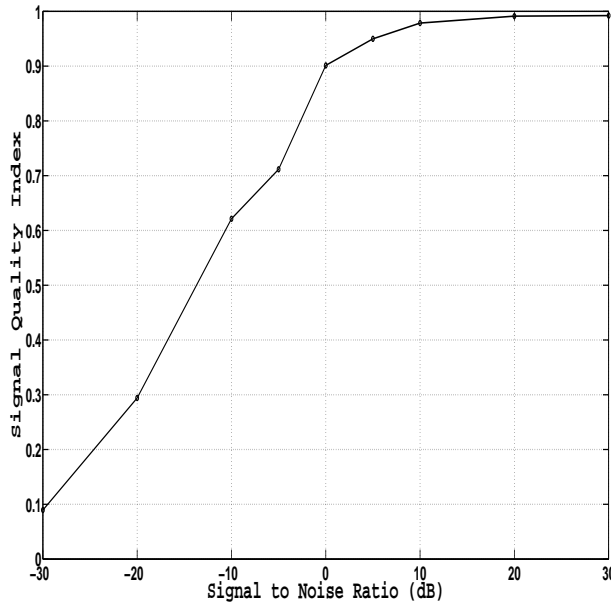


Fig. 4. Estimated SQI as a function of SNR.

TABLE II
SQI STATISTICS AS A FUNCTION OF HUMAN LABELING

	Good	Maybe	Bad
Mean	0.82	0.74	0.56
Standard Error	0.006	0.011	0.014

IV. DISCUSSION

The results show that the SQI estimated from Equation (3) is directly related to simulated SNR and human labeling. Most importantly, the estimated SQI is a monotonic function of either objective or subjective quality metrics. From a practical point of view, it seems that SQIs of above 0.8 and below 0.6 are sufficient to characterize a PLETH signal as “good” and “bad” quality respectively. From Figure 4 we can interpolate that SQI values of 0.8 and 0.6 are roughly equivalent to a simulated Gaussian noise SNR of about -2 and -10 dB respectively (if only uncorrelated noise is present).

While the algorithm for estimating signal quality described in this work shows promising results, it is also important to highlight its key assumptions. In particular, the algorithm assumes that the MCAF and its predictions are consistent and stable in the 30 second forecast window [1]. An unstable (or poor) prediction could be due to the MCAF algorithm rather than the signal quality in the target channel per se. Thus, the ability to detect instability on the MCAF prediction can be useful for improving the quality estimation. Another important assumption of the algorithm is the lack of correlation between noise in the target channel and the other measurement channels. Under some circumstances, such as in intense movement or when applying this algorithm to other channels, this assumption might not be valid. In

particular, when predicting SQI on ECG channels, removal of all other ECGs as inputs into the MCAF filter might be required (significantly degrading the MCAF’s predictive performance).

V. CONCLUSION

This work presented a new signal quality index for photoplethysmographs based on adaptive multichannel processing. The quality index was found to be monotonically related to both simulated signal-to-noise ratios and human quality perception of 1,313 waveforms. A quality index of 0.8 seems in general to be indicative of a “good” photoplethysmograph signal. Future recursive (i.e., online) implementation of this signal quality algorithm may also make it more attractive for a real-time applications.

VI. ACKNOWLEDGMENTS

This work was funded in part by the National Institute of Biomedical Imaging and Bioengineering and by the National Institute of General Medical Sciences, under NIH cooperative agreement U01-EB-008577 and NIH grant R01- EB001659. The authors wish to thank Yasemin Gokce for help with the human labeling and Daniel J. Scott for providing assistance with the manuscript.

REFERENCES

- [1] I. Silva, “PhysioNet 2010 Challenge : A Robust Multi-Channel Adaptive Filtering Approach to the Estimation of Physiological Recordings,” in *Computers in Cardiology*, September 2010, vol. 37, pp.313-316.
- [2] G. Moody, “The PhysioNet/Computing in Cardiology Challenge 2010: Mind the Gap,” in *Computers in Cardiology*, September 2010, vol.37 pp. 305-308.
- [3] M. Saeed, M. Villarroel, A. Reisner, G. Clifford, L. Lehman, G. Moody, T. Heldt, T. Kyaw, B. Moody, R. Mark, “Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC-II): A public-access intensive care unit database,” *Critical Care Medicine*, (In Press).
- [4] Fejzo Z, Lev-Ari L, “Adaptive Laguerre-Lattice Filters,” *IEEE Transactions on Signal Processing*, vol. 45(12), pp.3006-3016, 1997.
- [5] Haykin S, *Adaptive Filter Theory (Fourth Edition)*. Upper Sadle River, NJ: Prentice-Hall, 1990.