

# A Robust Fusion Model for Estimating Respiratory Rate from Photoplethysmography and Electrocardiography (Online Supplement)

Drew A. Birrenkott, Marco A.F. Pimentel, Peter J. Watkinson, and David A. Clifton

## S.1. ECG AND PPG RESPIRATORY MODULATIONS

The ECG and PPG are modulated by respiration in a number of ways, three of the most common being the waveform amplitude, frequency, and baseline wander as demonstrated in Fig. S.1.

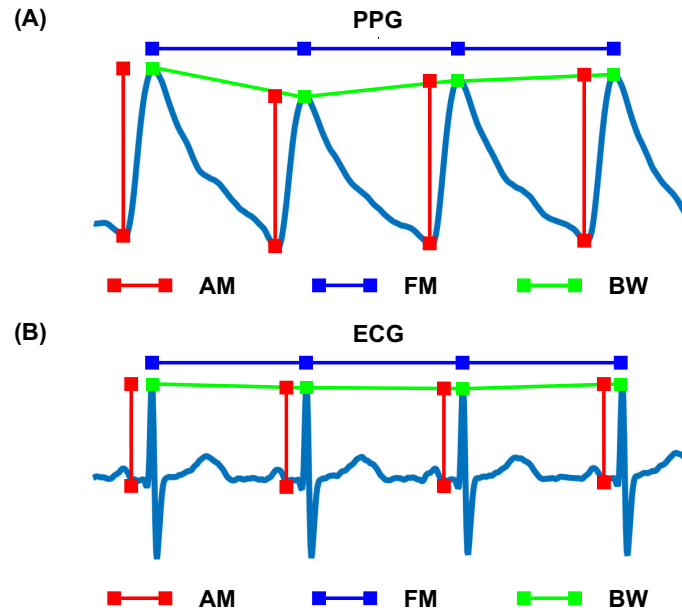


Fig. S.1. Common respiratory modulations present in both the (A) PPG and (B) ECG waveforms: amplitude modulation (AM), frequency modulation (FM), and baseline wander (BW).

## S.2. PHYSIOLOGICAL DESCRIPTION OF DATA SETS

Both CapnoBase ( $\mathcal{D}_c$ ) and ( $\mathcal{D}_m$ ) consist of subjects with similar mean HR and RR characteristics; however, the population age varies between the two data sets with  $\mathcal{D}_c$  consisting of a younger, more healthy population. The gender ratio was unavailable in the published information of the  $\mathcal{D}_c$  data set. The physiologic characteristics of each data set are shown in Table S.1.

TABLE S.1  
DEMOGRAPHIC AND PHYSIOLOGIC FEATURES OF THE SUBJECTS USED FROM THE CAPNOBASE AND MIMIC II DATA SETS.

Data Set	n	Age (Mean $\pm$ SD)	% Male	HR (Mean $\pm$ SD)	RR (Mean $\pm$ SD)
$\mathcal{D}_c$	42	20.9 $\pm$ 19.5		82.7 $\pm$ 20.8	14.8 $\pm$ 7.4
$\mathcal{D}_{c,1}$	32	21.5 $\pm$ 19.6		82.8 $\pm$ 21.3	14.7 $\pm$ 7.8
$\mathcal{D}_{c,2}$	10	19.1 $\pm$ 19.8		82.3 $\pm$ 20.0	15.1 $\pm$ 6.4
$\mathcal{D}_m$	53	64.8 $\pm$ 17.9	37.7%	88.9 $\pm$ 13.1	17.3 $\pm$ 3.0

### S.3. RR ESTIMATION ALGORITHM COMPARISON

Both the ARSpec and FFT-based algorithms have been used to estimate RR. The RR estimates for the two algorithms were compared in the  $\mathcal{D}_{c,1}$  data set by calculating the error between the estimate and the “gold standard” estimate,  $R_{ref}$ . The two algorithms had similar median and mean errors with ARSpec underestimating the actual RR more slightly more often than the FFT-based algorithm. These results are shown in Table S.2.

TABLE S.2  
COMPARISON OF MEDIAN (IQR) AND MEAN ERRORS ( $\pm$  S.D.) OF RR ESTIMATION METHODS AND  $R_{ref}$  IN  $\mathcal{D}_{c,1}$  WHERE ERROR IS CALCULATED AS:  
 $R_{est} - R_{ref}$ .

Estimation Method	Median Error (IQR)	Mean Error ( $\pm$ SD)
ARSpec (Shah <i>et al.</i> 2015)	-0.40 (-1.40 - 0. 17)	-0.85 $\pm$ 6.96
FFT Method (Karlen <i>et al.</i> 2013)	-0.15 (-0.89 - 0.75)	-0.37 $\pm$ 7.32

#### S.4. LOGISTIC FUNCTION BOUNDING OF FUSION RQIS

The RQI fusion models,  $\tilde{Q}_F$ , described in section II.E were trained to estimate the RR error  $E$ , and this results in windows with high values for the individual RQIs being mapped onto low output values from the RQI-fusion regressor (corresponding to those windows having low associated error  $E$  with a standard RR-estimation method). Conversely, for windows with low values for the individual RQIs (representing a window of low-quality modulation), the RQI-fusion regressor outputs high values (corresponding to those windows having high associated error with a standard RR-estimation method). The resultant fused RQIs were unbounded,  $y \in \mathbb{R}$ . Therefore, we transformed the resultant fused RQIs using the logistic function:

$$Q_{F,i} = \frac{1}{1 + e^{(\tilde{Q}_{F,i} - \bar{Q}_{F,i})}} \quad (1)$$

for  $i = 1 \dots 3$ , where  $i$  represents each fusion model,  $\tilde{Q}_{F,i}$  is the vector of all fusion RQIs,  $\bar{Q}_{F,i}$  is the mean of the fusion RQIs from the  $\mathcal{D}_{c,1}$ , and  $Q_{F,i} \in [0 \ 1]$  is the bounded fused RQI. In using this sign-inverted transformation, the transformed RQI,  $Q_{F,i}$ , takes high values when the four input RQIs take high values, and conversely takes low values when the input RQIs take low values. Thus the final outcome of our procedure is a single fused RQI (for each modulation, for each window), that has been transformed to the range  $[0 \ 1]$  and where higher values indicate higher-quality modulation.

### S.5. CORRELATION ANALYSIS BETWEEN $Q_{S,1}$ , $Q_{S,2}$ , AND $Q_{S,3}$

The aggregate performance of the individual RQIs for  $E_{50}$  and  $\rho$  were similar. However, the rank order correlation of these values indicates that the individual RQIs did not perform identically (Table S.3).

TABLE S.3

SPEARMAN'S RANK ORDER CORRELATION COEFFICIENT,  $\rho$ , FOR THE INDIVIDUAL RQIS,  $Q_{S,k}$  FOR  $k = 1 \dots 3$  FOR  $\mathcal{D}_{c,1}$ ,  $\mathcal{D}_{c,2}$ , AND  $\mathcal{D}_m$

	$\mathcal{D}_{c,1}$			$\mathcal{D}_{c,2}$			$\mathcal{D}_m$		
	$Q_{s,1}$	$Q_{S,2}$	$Q_{S,3}$	$Q_{S,1}$	$Q_{S,2}$	$Q_{S,3}$	$Q_{S,1}$	$Q_{S,2}$	$Q_{S,3}$
$Q_{S,1}$	1.00			1.00			1.00		
$Q_{S,2}$	0.79	1.00		0.83	1.00		0.71	1.00	
$Q_{S,3}$	0.85	0.79	1.00	0.87	0.86	1.00	0.77	0.70	1.00

### S.6. PERFORMANCE OF RQIS FOR INDIVIDUAL MODULATIONS AND WINDOWS IN $\mathcal{D}_{c,1}$

Fig. S.2 is an identical analysis on the training and validation data set,  $\mathcal{D}_{c,1}$  as was conducted on both of the test data sets,  $\mathcal{D}_{c,2}$  and  $\mathcal{D}_m$  in Fig. 2.

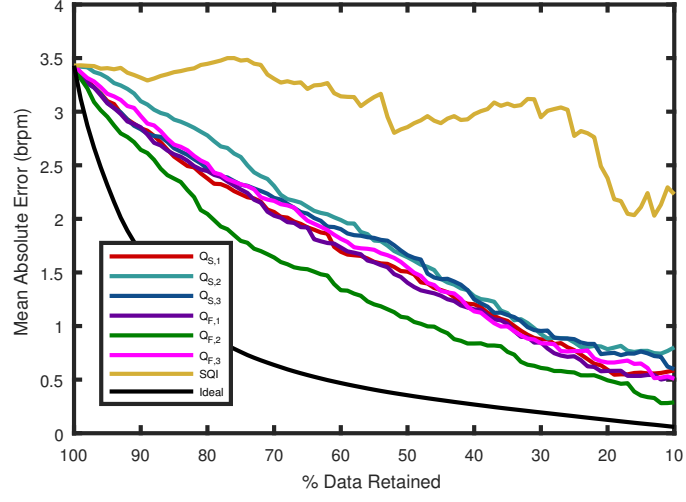


Fig. S.2. MAE of  $Q_{S,k}$  (for  $k = 1 \dots 3$ ),  $Q_{F,i}$ , and controls on the  $\mathcal{D}_{c,1}$  data set as windows are discarded based on comparison of RQIs to a threshold  $V$ . Note that the horizontal axis shows the percentage of data retained, which decreases as the value of  $V$  increases.

### S.7. FUSION MODEL HYPERPARAMETERS

The determination of the hyperparameters for LR and SVR models is described in section II.E. In the LR model, the optimal value for  $\alpha$  was 0 (Fig. S.3). In the SVR model, the optimal values for  $C$  was 8 and for  $\gamma$  was 0.5 (Fig. S.4 and Fig. S.5).

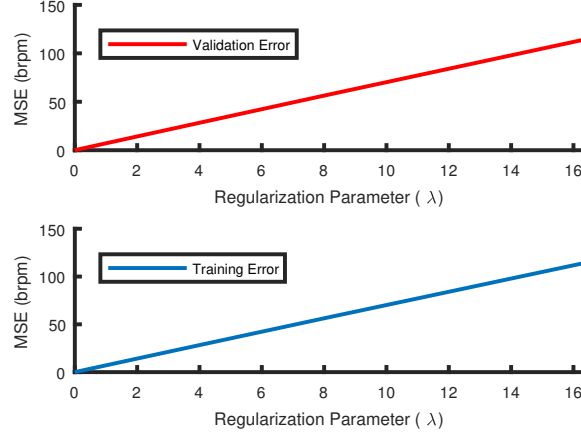


Fig. S.3. Average mean squared error (MSE) values for linear regression fusion model. The optimal regularization parameter for the model is  $\lambda = 0$ .

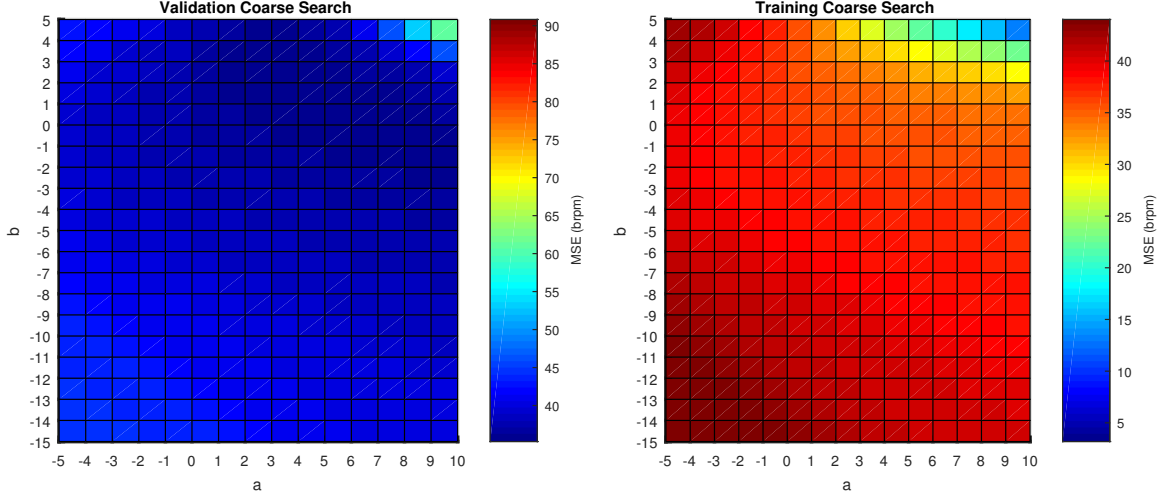


Fig. S.4. Average MSE values for coarse grid search on validation (left panel) and training (right panel) sets for SVR fusion model.  $C = 2^a$  and  $\gamma = 2^b$ .

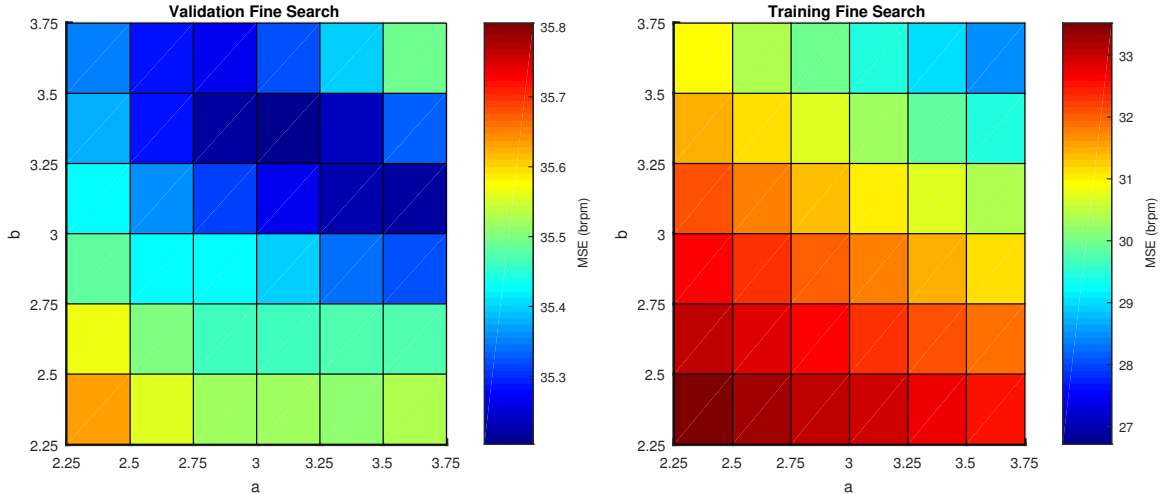


Fig. S.5. Average MSE values for fine grid search on validation (left panel) and training (right panel) sets for SVR fusion model.  $C = 2^a$  and  $\gamma = 2^b$  with optimal values for  $C = 8$  ( $a = 3$ ) and  $\gamma = 9.5$  ( $b = 3.25$ ).

### S.8. COMPARATIVE PERFORMANCE OF $Q_{F,1}$ TO $Q_{S,k}$ FOR $k = 1 \text{ dots } 3$ IN $\mathcal{D}_{c,1}$ , $\mathcal{D}_{c,2}$ , AND $\mathcal{D}_m$

As described in section III.B, the fusion model,  $Q_{F,1}$  is a more robust model than any of the individual RQIs. This is demonstrated in Tables S.4, S.5, and S.6 which show the MAE of all three methods in all three data sets when either 50% or 75% of the individual modulations and windows have been discarded. These results show that  $Q_{F,1}$  is either the best or second-best performing RQI method in all instances when considering both the MAE and the standard deviation of the MAE. This indicates that  $Q_{F,1}$  has both high accuracy and precision across all data sets compared to the individual RQIs indicating high robustness for the fusion method.

TABLE S.4  
MAE AND SD FOR  $Q_{S,1\dots3}$  AND  $Q_{F,1}$  IN  $\mathcal{D}_{c,1}$  WHEN 50% OR 75% OF WINDOWS HAVE BEEN DISCARDED BASED ON RQI VALUE.

	MAE 50% Discard	MAE 75% Discard
$Q_{S,1}$	$1.51 \pm 0.78$	$0.78 \pm 2.25$
$Q_{S,2}$	$1.64 \pm 0.85$	$0.85 \pm 2.16$
$Q_{S,3}$	$1.66 \pm 0.88$	$0.88 \pm 2.49$
$Q_{F,1}$	$1.40 \pm 0.72$	$0.72 \pm 1.94$

TABLE S.5  
MAE AND SD FOR  $Q_{S,1\dots3}$  AND  $Q_{F,1}$  IN  $\mathcal{D}_{c,2}$  WHEN 50% OR 75% OF WINDOWS HAVE BEEN DISCARDED BASED ON RQI VALUE.

	MAE 50% Discard	MAE 75% Discard
$Q_{S,1}$	$0.68 \pm 1.17$	$0.59 \pm 1.05$
$Q_{S,2}$	$0.81 \pm 1.96$	$0.52 \pm 1.36$
$Q_{S,3}$	$0.58 \pm 0.90$	$0.48 \pm 0.92$
$Q_{F,1}$	$0.63 \pm 1.12$	$0.52 \pm 0.98$

TABLE S.6  
MAE AND SD FOR  $Q_{S,1\dots3}$  AND  $Q_{F,1}$  IN  $\mathcal{D}_m$  WHEN 50% OR 75% OF WINDOWS HAVE BEEN DISCARDED BASED ON RQI VALUE.

	MAE 50% Discard	MAE 75% Discard
$Q_{S,1}$	$3.85 \pm 4.78$	$2.92 \pm 4.55$
$Q_{S,2}$	$3.86 \pm 5.19$	$2.95 \pm 4.89$
$Q_{S,3}$	$4.08 \pm 5.01$	$3.49 \pm 4.91$
$Q_{F,1}$	$3.82 \pm 4.85$	$2.91 \pm 4.60$



### S.9. SELECTION OF EXEMPLAR VALUES FOR THRESHOLD $T$

In addition to the selection of the exemplar threshold used for RQIFusion (here termed  $T_{HT}$ ), we chose two additional exemplar values of threshold  $T$ : a low threshold ( $T_{LT}$ ) and an ultra-high threshold ( $T_{UHT}$ ).  $T_{LT}$  was set to the highest possible value of  $T$  such that every window had a least one modulation with RQI  $Q_{F,1,j} > T_{LT}$ ,  $V_j \in [1 \dots 6]$ . An RR estimation was therefore obtained for every window in the  $\mathcal{D}_{c,1}$  data set.  $T_{UHT}$  was set to compare the results of the proposed algorithm with the those of SmartFusion. The value of  $T_{UHT}$  was therefore set such that the proposed method had the same approximate data-discard rate (34.0%) as SmartFusion (34.3%) in the  $\mathcal{D}_{c,1}$  data set. The above resulted in  $T_{LT} = 0.11$ ,  $T_{HT} = 0.44$  (as described), and  $T_{UHT} = 0.84$ , as shown in Fig. S.6. The figure re-plots the MAE shown previously in Fig. 3, against the values of  $T$ , illustrating the relationship of the  $T_{LT}$ ,  $T_{HT}$ , and  $T_{UHT}$  values percent of windows for which no RR estimate is available using the proposed method. The performance of all the the exemplar thesholds comapred to the existing methods for all three data sets is shown in Table S.7.

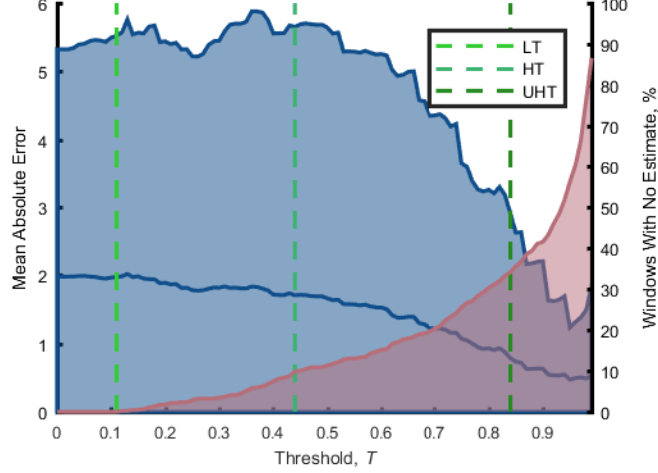


Fig. S.6. (median and s.d.) MAE (blue, vertical axis) and the number of windows with no estimate (red, vertical axis) for  $\mathcal{D}_{c,1}$  vs. increasing values of threshold  $T$  (horizontal axis). The dashed lines represent the values of  $T_{LT}$  (light green, leftmost),  $T_{HT}$  (mid-green, center-most), and  $T_{UHT}$  (dark green, rightmost).

TABLE S.7

PERFORMANCE OF EXEMPLAR THRESHOLD MODELS (MEDIAN, MAE, AND % DISCARD) COMPARED TO EXISTING FUSION METHODS FOR ALL THREE DATA SETS,  $\mathcal{D}_{c,1}$ ,  $\mathcal{D}_{c,2}$ , AND  $\mathcal{D}_m$

	Median (Q1-Q2)	$\mathcal{D}_{c,1}$ MAE ( $\pm$ SD)	% Dis.	Median (IQR)	$\mathcal{D}_{c,2}$ MAE ( $\pm$ SD)	% Dis.	Median (IQR)	$\mathcal{D}_m$ MAE ( $\pm$ SD)	% Dis.
<b>Avg.</b>	1.34 (0.63 - 3.26)	2.61 $\pm$ 3.30	0.0%	0.81 (0.39 - 1.39)	1.06 $\pm$ 0.93	0.0%	3.54 (1.72 - 5.88)	4.04 $\pm$ 2.90	0.0%
<b>SF</b>	0.62 (0.26 - 1.18)	1.23 $\pm$ 2.42	34.0%	0.46 (0.23 - 0.92)	0.68 $\pm$ 0.61	18.0%	2.83 (0.66 - 8.51)	4.61 $\pm$ 4.50	58.8%
<b>Mod. SF</b>	0.82 (0.41 - 1.31)	1.00 $\pm$ 0.92	46.0%	0.65 (0.31 - 1.08)	0.76 $\pm$ 0.56	26.7%	2.20 (1.21 - 4.51)	3.33 $\pm$ 3.00	73.0%
<b><math>T_{LT}</math></b>	0.77 (0.31 - 1.75)	1.98 $\pm$ 3.55	0.0%	0.63 (0.31 - 1.20)	0.86 $\pm$ 0.72	0.0%	3.09 (1.23 - 5.86)	3.97 $\pm$ 3.45	1.2%
<b><math>T_{HT}</math> (RQIFusion)</b>	0.49 (0.24 - 1.08)	1.71 $\pm$ 3.94	10.0%	0.43 (0.25 - 0.77)	0.71 $\pm$ 0.89	1.3%	0.72 (0.27 - 4.91)	3.12 $\pm$ 4.39	23.2%
<b><math>T_{UHT}</math></b>	0.34 (0.17 - 0.62)	0.79 $\pm$ 2.13	34.0%	0.34 (0.21 - 0.57)	0.53 $\pm$ 0.82	14.7%	0.41 (0.15 - 1.07)	2.38 $\pm$ 5.09	66.3%

### S.10. COMPARISON OF RQIFUSION TO SMARTFUSION USING PPG OR ECG ONLY

Because SmartFusion only uses PPG waveforms to extract respirator modulations, we compared RQIFusion to SmartFusion when RQIFusion only used either PPG or ECG modulations. The threshold,  $T$  was set to have the identical discard percentage as SmartFusion for the given setting and the median and MAE results are presented in Table S.8 and show that RQIFusion outperforms SmartFusion in all instances.<sup>1</sup>

TABLE S.8

MEDIAN (IQR), MEAN ABSOLUTE ERROR ( $\pm$  S.D.), AND % OF WINDOWS DISCARDED FOR SMARTFUSION AND THE CORRESPONDING THRESHOLD VALUE  $T$  WHEN % DISCARD VALUES ARE SET EQUAL AND EITHER PPG MODULATIONS OR ECG MODULATIONS ARE CONSIDERED IN  $\mathcal{D}_{c,2}$  AND  $\mathcal{D}_m$ .

Data Set	Modality	Method	Median (IQR)	MAE ( $\pm$ SD)	% Dis.
$\mathcal{D}_{c,2}$	PPG	SF (PPG only)	0.46 (0.23 - 0.92)	$0.68 \pm 0.61$	18.0%
$\mathcal{D}_{c,2}$	PPG	$T = 0.69$	0.37 (0.17 - 0.67)	$0.52 \pm 0.62$	18.0%
$\mathcal{D}_{c,2}$	ECG	SF (ECG only)	0.70 (0.33 - 1.24)	$0.88 \pm 0.78$	24.0%
$\mathcal{D}_{c,2}$	ECG	$T = 0.70$	0.47 (0.21 - 0.69)	$0.68 \pm 1.29$	24.0%
$\mathcal{D}_m$	PPG	SF (PPG only)	2.83 (0.66 - 8.51)	$4.61 \pm 4.50$	58.8%
$\mathcal{D}_m$	PPG	$T = 0.68$	0.43 (0.18 - 1.35)	$2.28 \pm 3.83$	58.6%
$\mathcal{D}_m$	ECG	SF (ECG only)	1.62 (0.54 - 4.77)	$3.45 \pm 4.13$	53.9%
$\mathcal{D}_m$	ECG	$T = 0.41$	0.65 (0.26 - 4.13)	$3.10 \pm 5.10$	53.6%

<sup>1</sup>For the ECG analysis, SmartFusion was applied to the three ECG modulations alone. While this was not done in the original SmartFusion publication, Karlen *et al.* (2013), this was done to allow for a consistent ECG-alone comparison.

### S.11. PERFORMANCE OF RQIFUSION ON REMAINING INDIVIDUAL PATIENTS FROM $\mathcal{D}_{c,2}$

Fig. 4 shows the performance of RQI fusion in tracking RR over time for 2 of the 10 patients from  $\mathcal{D}_{c,2}$ , Fig. S.7 shows the results for the remaining 8 patients.

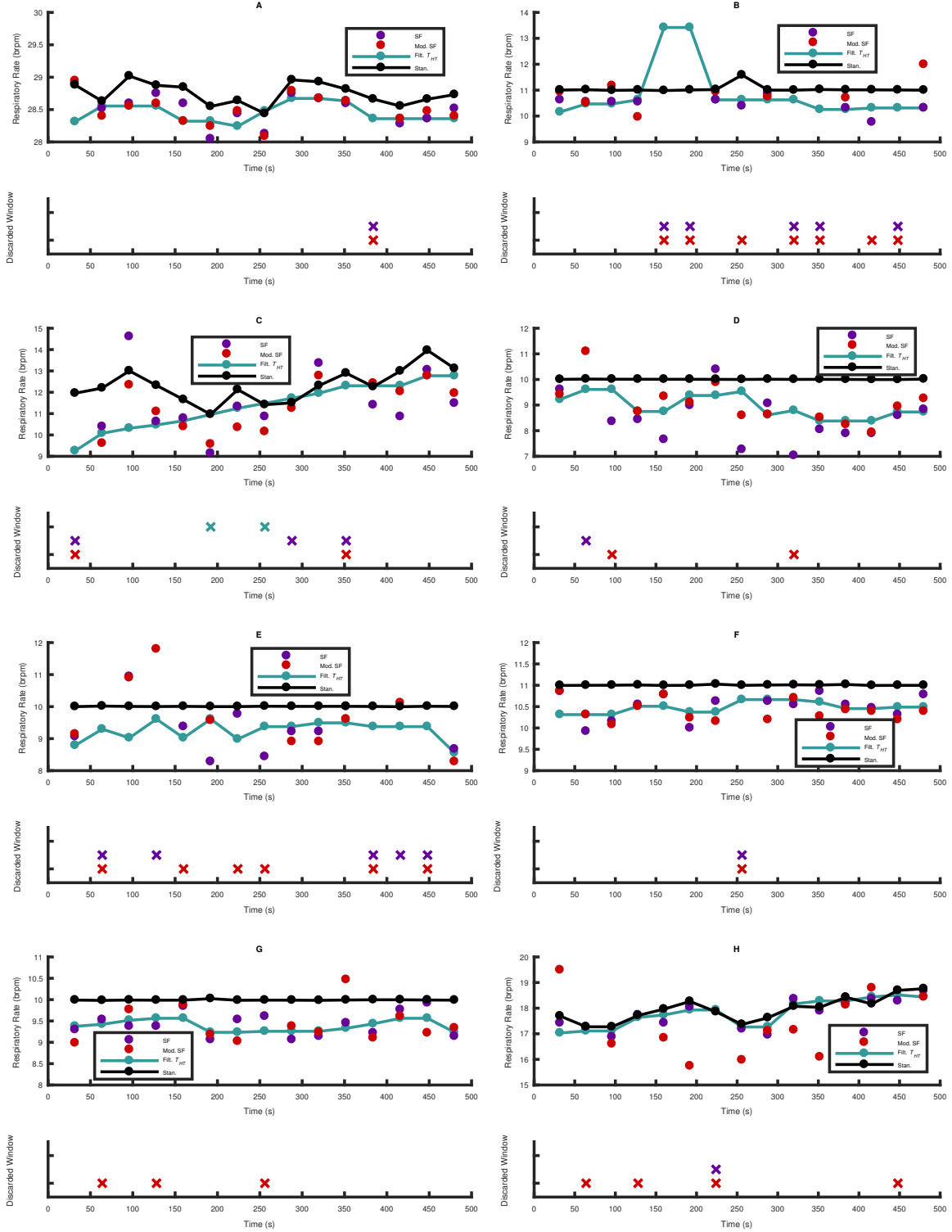


Fig. S.7. (Upper row) Time-series of RR estimates for modified SmartFusion, SmartFusion, the HT-based proposed method, and the  $R_{ref}$  reference values for subjects in data set  $\mathcal{D}_{c,2}$  not shown in figure 5: (A) subject 15, (B) subject 123, (C) subject 128, (D) subject 148, (E) subject 149, (F) subject 150, (G) subject 313, and (H) subject 322. (Lower row) Corresponding plots showing those windows for which each method did not produce an RR estimate.

## S.12. LIMITATIONS OF PRESENT STUDY

### A. *Limitations of this Study*

Further validation of the proposed method is still required. The current results demonstrate the efficacy of the method with subsets of the CapnoBase and MIMIC II data sets, although (in keeping with the literature) we used records from only 63 independent subjects, with 8 minutes of data for each subject. Furthermore, these subjects do not necessarily represent a true clinical setting: the CapnoBase subjects were generally healthy and anesthetized, and the subset of MIMIC II subjects was selected for the presence of high-fidelity signals. However, large-scale clinical validation remains a challenge because a ground truth value for RR is required. While methods such as capnography and IP can be used as a proxy for this, the quality of the RR estimates from these methods also relies on the presence of high signal fidelity. This means that without knowledge of the quality of the capnography or IP signal seemingly high errors in RR estimation from PPG or ECG may be incorrectly attributed to failures in the RR-estimation method, whereas the actual source of the error may actually be due to an imperfect reference. The current strategy to address this required hand-scoring of reference signals by the authors; however, this approach is infeasible for a large-scale clinical validation.

Furthermore, the current method was validated with data from subjects who were either anesthetized or otherwise immobile. The actual clinical usefulness of RR estimation from the ECG or PPG with such patients is minimal, however, as immobile patients in the hospital are usually those who are most likely to be monitored using dedicated techniques such as capnography and IP. Thus one of the most promising avenues for further development of the proposed method is for use in general and ambulatory wards, in rehabilitation centers, at home following discharge from hospital, and in using wearable sensors for the general population. Data from such patients are likely to have frequently-corrupted data due to motion artifact and other factors.<sup>2</sup> For monitoring mobile patients, the application of SQIs (such as the F1 score) could be used as a pre-processing step to remove data that are clearly artifactual before use of RQIs, as described in this paper, to identify the presence or absence of respiratory information in the resulting (non-artifactual data).

<sup>2</sup>G. D. Clifford and D. Clifton, "Wireless technology in disease management and medicine," *Annu Rev Med*, vol. 63, pp. 47992, 2012.