Cuffless Blood Pressure Estimation during Moderate- and Heavy-Intensity Exercise – Supplementary Material

Cederick Landry, Member, IEEE, Eric T. Hedge, Richard L. Hughson, Sean D. Peterson, and Arash Arami*, Member, IEEE

Main paper is published in IEEE Journal of Biomedical and Health Informatics (2022).

DOI: 10.1109/JBHI.2022.3207947

The main manuscript presents the results for the L-H exercise. Herein the analogous results for the L-M exercise are presented. In this section, one participant was removed from the analysis (out of 12) due to a poor PPG signal biasing the results. The PPG signal was found to have small signal to noise ratio, making it difficult to extract the PPG peak.

The ranges of MAP and ΔMAP covered by L-H are shown in Fig. S.1. The broader distribution in MAP than ΔMAP is explained by the baseline MAP which is different from one participant to another. ΔMAP reflects how MAP changes during L-H, where ΔMAP = 0 mmHg indicates the average MAP during the experiment.

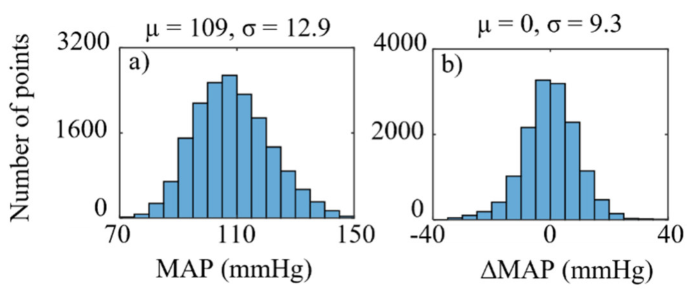

Fig. S.1 - Distribution of a) MAP and b) ΔMAP for all participants during L-M.

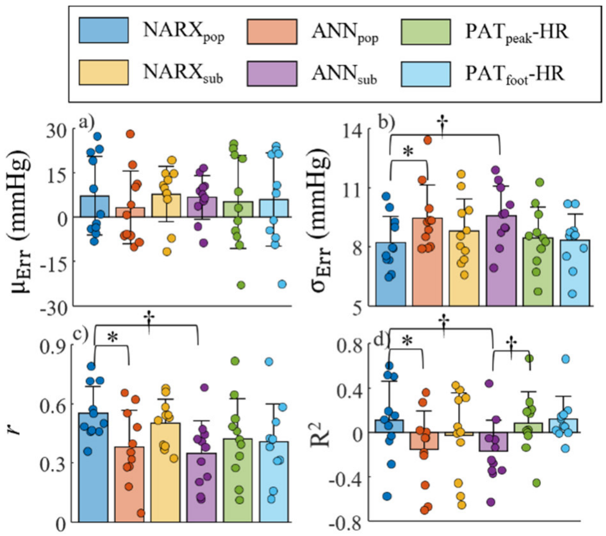

Fig. S.2 - Comparison of a) mean error, b) SD of the error, c) Pearson correlation coefficient and d) coefficient of determination between different model estimates and the BP measurements. Bars represent the mean of all subjects and the error bars show ± SD for the L-M exercise. Each data point represents one participant. Legend: Wilcoxon with (*) p<0.05/15 and (†) p<0.01.

The mean and SD of μ_{Err}, σ_{Err}, r, and R^2 during the L-M exercise computed across the subjects are shown in Fig. S.2 for the different models. The SD of μ_{Err} is relatively large, showing that all BP models can be largely biased for some participants. The results exhibit differences in mean σ_{Err} between the models, ranging from 8.19 mmHg for $NARX_{pop}$ (best) to 9.58 mmHg for ANN_{sub} (worst). As a reference, the MAP SD across the participants is 9.1±1.6 mmHg (mean±SD) during the L-M exercise. Due to the distribution of results, only the $NARX_{pop}$ σ_{Err} is found to be statistically lower than ANN_{pop} ($p<0.05/10$), and trends toward a statistical difference with ANN_{sub} ($p<0.01$). The average Pearson correlation coefficient between the estimated and measured MAP is higher for $NARX_{pop}$ and statistically different from ANN_{pop}(ANN_{sub}) ($p<0.05/10$). Similarly, R^2 is the largest for both $NARX_{pop}$ and statistically different from ANN_{pop}(ANN_{sub}) ($p<0.05/10$) and PAT-HR is statistically different from ANN_{sub} ($p<0.01$).

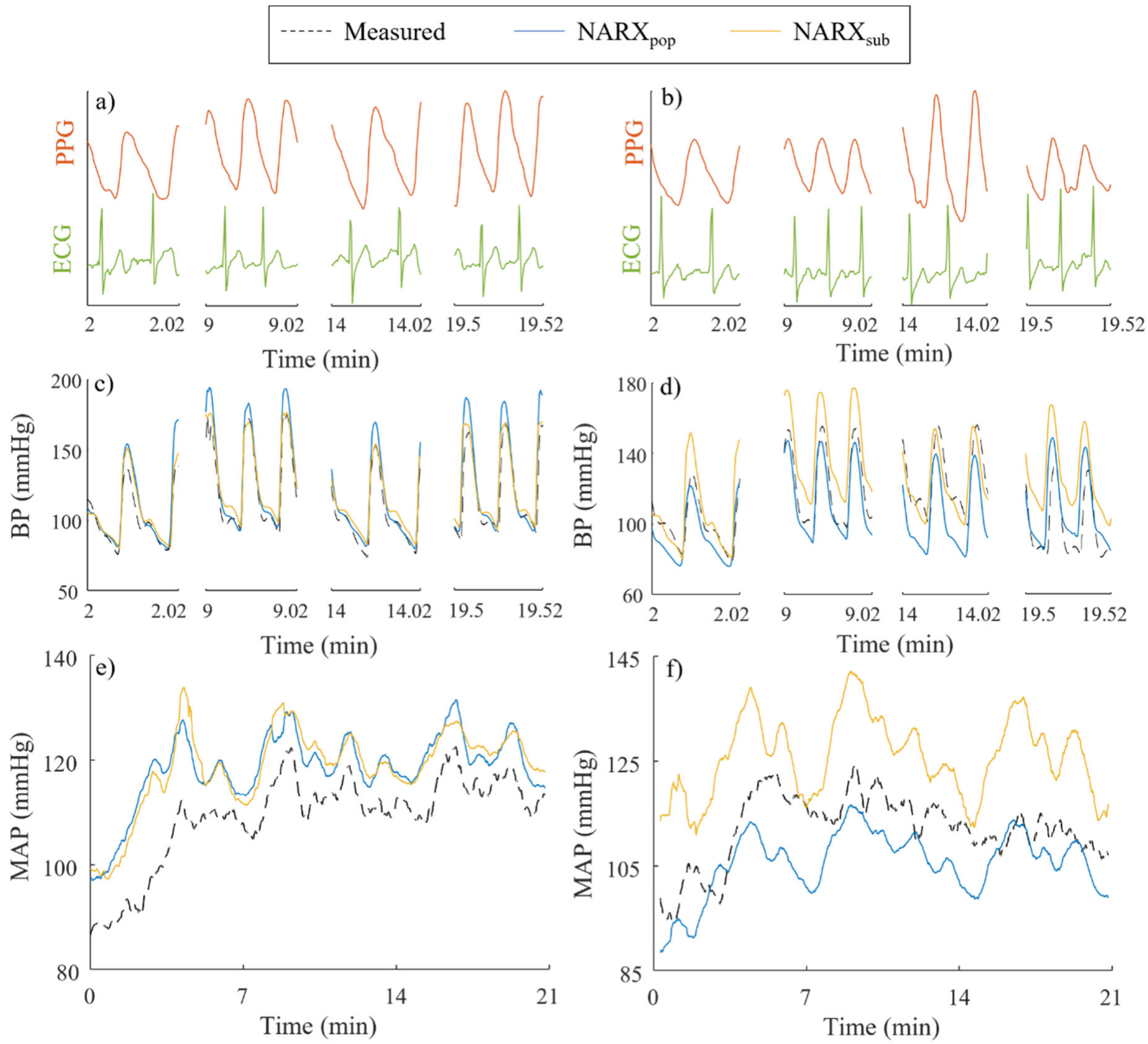

Fig. S.3 - a) - b) PPG and ECG waveforms and c) - f) comparison of actual versus estimated BP waveforms and MAP (after applying a moving average). The results are shown for the subject with the lowest, a), c), and e), and highest, b), d), and f), SD of the error for the L-M exercise.

The estimated BP time series from $NARX_{pop}$ and $NARX_{sub}$ are shown for the subjects with the lowest and highest σ_{Err} in Fig. S.3. In Fig. S.3.a-d, the PPG, ECG, and estimated BP waveforms are shown at rest (2 min), during the first PRBS moderate-intensity (90%VT) section (9 min), during the low-intensity (25 W) section between the two PRBS (14 min), and at the last moderate-intensity (90%VT) section (19.5 min). In Fig. S.3.c-d, MAP is shown, as a function of time, for the complete 21 minutes, including the initial resting portion and the complete L-M exercise.

The $NARX_{pop}$ error (with removed bias) evolution over the 21-minute data collection period is presented for all participants in Fig. S.4. The red line is the mean of all participants. It can be observed that the model in the majority of subjects underestimates the change in MAP up to the end of the first PRBS (13 min) and error slowly increases throughout the exercise, which is consistent with qualitative observations in Fig. S.3.

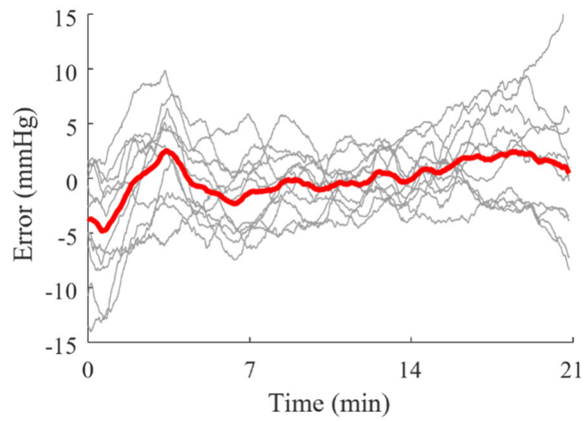

Fig. S.4 - Filtered BP error for the 11 subjects for the L-M exercise. Grey lines represent individual subject data and the red line is the mean of all subjects.

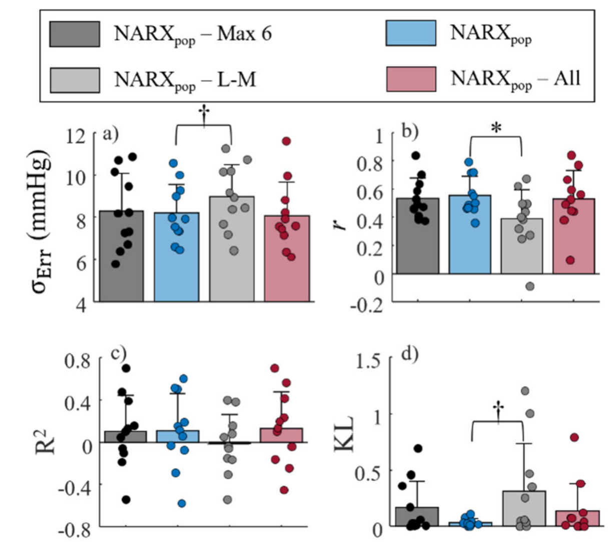

Fig. S.5 - Comparison of a) SD of the error, b) Pearson correlation coefficient, c) coefficient of determination, and d) KL divergence between different model estimates and the BP measurements distributions. Bars represent the mean of all subjects and the error bars show ± SD for the L-M exercise. Each data point represents one participant. Legend: Wilcoxon with (*) $p<0.05/6$ and (†) $p<0.05$.

The means and SDs of σ_{Err}, r, R^2, and KL-divergence during the L-M exercise computed across the 11 subjects are shown in Fig. S.5 for $NARX_{pop}$ trained on different data . In general, training the model on only the L-M exercise ($NARX_{pop}$ – L-M) leads to poorer outcomes, with statistical difference observed between r of $NARX_{pop}$. Training on the Max exercises ($NARX_{pop}$) resulted in higher r, and significantly lower KL-divergence than training the model on L-M exercises. On average, training on all exercises ($NARX_{pop}$ – All) leads to larger R^2 and smaller σ_{Err} compared to the other models.

The capability of predicting changes in MAP by $NARX_{pop}$ is shown in Fig. S.6 for the different training datasets. As shown by the p-values obtained from the Wilcoxon tests, using the data of all the exercises as training data leads to lower MAE in the small range of ΔMAP without affecting the model's capability to estimate BP at large ΔMAP. The model trained on the Max exercises is significantly better at estimating large positive ΔMAP than the training only on L-M. There were no significant differences between training on six or 11 participants Max exercises when comparing every ΔMAP bin, so nothing appears on the figure.

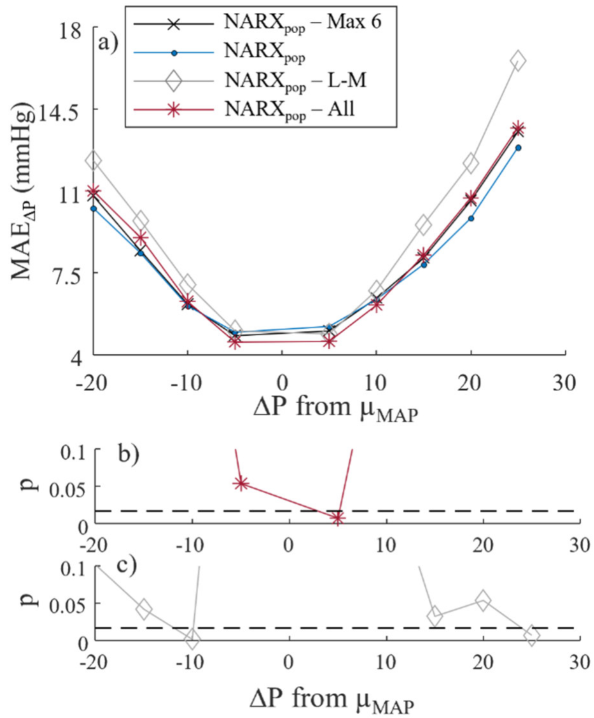

Fig. S.6 - a) Mean of $MAE_{\Delta P}$ for all participants against the distribution of ΔMAP from μMAP measured throughout the L-M exercise. Wilcoxon tests' p-value for b) $NARX_{pop}$ vs $NARX_{pop}$ – All, c) $NARX_{pop}$ vs $NARX_{pop}$ – L-M, and d) $NARX_{pop}$ vs $NARX_{pop}$ – Max6. Symbols represent the p-values (or $MAE_{\Delta P}$ value) and the location on the abscissa indicates the bin over which the statistical test was performed.

The means and SDs of the σ_{Err}, r, R^2, $r_{err\text{-}PPG}$, and $r_{err\text{-}ECG}$ during the L-M exercise computed across the subjects are shown in Fig. S.7 for $NARX_{pop}$, $NARX_{sub}$ – L-M (Train), $NARX_{sub}$ – All (Train), and $NARX_{corr}$. The results indicate that when a model can extract more information from the PPG and ECG signals (lower $r_{err\text{-}PPG}$ and $r_{err\text{-}ECG}$), it is better at estimating BP (lower σ_{Err} and higher r). It can also be observed that the employed NARX architectures are not able to extract all the information from the PPG and ECG even when the exercise data are included in the training set, though this helps the estimation for the moderate exercise case ($NARX_{sub}$ – All (Train)).

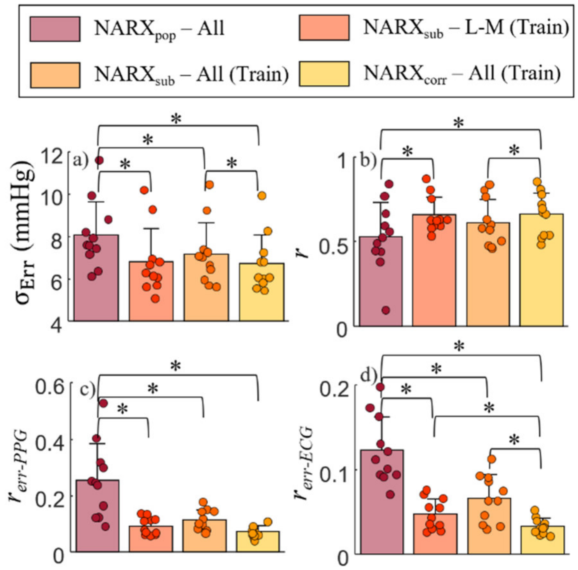

Fig. S.7 - Comparison of a) SD of the error, and b) Pearson correlation coefficient between different model estimates and the BP measurements, and the maximum cross-correlation between c) BP and PPG, and d) BP and ECG signals. Bars represent the mean of all subjects and the error bars show ± SD for the L-M exercise. Each data point represents one participant. Legend: (*) Wilcoxon with p<0.05/6.

The mean of the participants' ΔMAP during the L-M exercise is shown in Fig. S.8.a. It can be seen that the measured ΔMAP during the second PRBS is lower than the first one, but the difference is much smaller compared to the L-H exercise. Fig. S.8.b shows HR, PPG_{peak} and 1/PAT, which are all features that keep increasing throughout the exercise. $NARX_{pop}$ – All – Cal shows a better tracking of the change in ΔMAP between the two PRBS. As a results, the μ_{Err} decreases to 1.4 ± 3.6 mmHg and σ_{Err} to 7.5 ± 1.6 mmHg, as opposed to μ_{Err} = 0.7 ± 11.9 mmHg and σ_{Err} = 8.1 ± 1.6 mmHg for $NARX_{pop}$ – All.

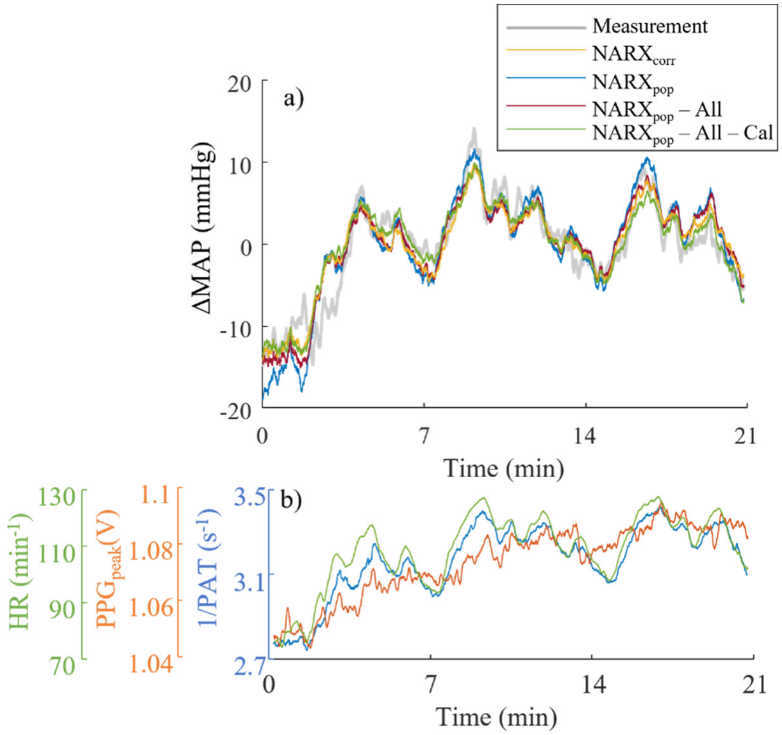

Fig. S.8 – Mean of the 11 participants a) measured and estimated ΔMAP and b) common features that correlate with BP for the L-M exercise.