Detecting Stress Using Wearables

Matty Pahren, Scott Heng On, Ethan Shen

1 Introduction

Affect recognition seeks to detect a person's affective state. One of these states is stress, and long-term stress is known to have severe implications on human health and well-being [1]. With the rapid developments in wearable health technology, tracking various body metrics can allow to determine one's affective state. This study uses the Wearable Stress and Affect Detection (WESAD) dataset that contains both wrist and chest collected data, which was introduced in Schmidt et al. [2], and achieves classification accuracies of 86.2% on wrist data and 90.8% on all data (both chest and wrist) for the binary classification problem (stress vs. amusement). In conducting this analysis, we aim to determine whether sensor data are useful in predicting stress, and if so, which predictors are most significant when discriminating between stress and amusement. Furthermore, we seek to understand which types of sensor data are most useful in predicting stress, alone or in combination, and to determine if stress can be detected only using the wrist-worn wearable, which is more convenient to wear than a chest-worn device. Finally, we quantify the heterogeneity across individuals in their responses to stress.

2 Literature Review

Stress can impact heart rate, blood pressure, and skin temperature [3]. When a threat is presented, it only takes the brain .9 milliseconds, or 1111 Hz, to shift a body into a full stress response [4]. In that time, the brain filters information relevant to our personal safety, redirects energy to only areas that are biologically relevant for survival, and releases stress hormones. While stress can be useful in some situations in that it activates fight-or-flight response and can help us during a dangerous situation, chronic stress can have many negative long-term impacts, even reducing brain mass in some cases.

However, stress is not the only emotion that impacts us physically. In fact, many of our physiological measures change based on our emotional state. L. Shu et. al. [5] examine how anger, anxiety, embarrassment, fear, amusement, happiness, and joy have different effects on our cardiovascular, electrodermal, respiratory, and electroencephalographic measures. Specifically, they saw that when patients are amused, heart rate becomes more variable and electrodermal activity, respiration rate and brain activity all increase.

With more advanced technology these days, studies have shown that changes in emotion can be effectively measured by sensors of physiological data [6]. This study found that emotional valence and arousal could all be reliably estimated by using low-cost sensors used to collect data on electroencephalography (EEG), galvanic skin response (GSR), and electromyographic signal (EMG) data.

There has also been research done on the appropriate time intervals to use when measuring different body measures such as motion, respiration, electrodermal activity, and heart rate. Different studies use various sampling rates of the raw signals and window sizes for feature engineering, and it appears that there is no consensus on the best-suited way to choose these values. Arif and Kattan [7] used a five second rolling window to analyze ACC data, and Uy et. al. [8] also used a five second rolling window for measuring Galvanic Skin Response, Blood Volume Pulse, and Respiratory Variability. Since we found multiple sources using a five-second window for predictors similar to the ones in our dataset, we decided to use this approach as well.

3 Data

The WESAD dataset [2] contains 63 million samples of raw physiological and motion signal data collected from 15 test subjects (12 male, 3 female) with ages between 24-35. The data was collected using 2 devices: a chest-worn

device (RespiBAN) and a wrist-worn device (Empatica E4). The RespiBAN chest data includes electrocardiography (ECG) data in mV, electrodermal activity (EDA) in microseconds, electromyography (EMG) data in mV, skin temperature (TEMP) in °C, displacement of the thorax for males and the abdomen for females induced by inhaling or exhaling (RESP) as a percentage, and 3-axis accelerometer (ACC) data in g - the acceleration of gravity. All RespiBAN data were recorded at 700 Hz. The E4 wrist data collected includes blood volume pulse (BVP) data (64Hz), ACC data in units of 1/64g (32Hz), EDA in microseconds (4Hz), and TEMP in °C (4Hz). Each sample also references a label corresponding to the study protocol condition, with each label referring to a different condition (1-baseline, 2-stress, 3-amusement, 4-meditation). These conditions were recorded by having participants exposed to sources of stimulation that would elicit the desired condition. The labels 0, 5, 6, 7 indicate that the data should be ignored. The dataset also included information about each subject, such as height, weight and age, as well as personal responses to PANAS and SSSQ questionnaires. Considering the nature of the WESAD dataset and the scope of our study, we decided to filter out samples that did not represent stress or amusement.

Figure 1.1 and 1.2 show plots of the EDA signal data by participant and by condition showing different values when the participant experienced different conditions, and this difference is captured both in the chest and wrist data. Since we only took samples for 2 conditions, we could desirably observe a relative equal distribution of samples in both groups. However, raw signal data only provides limited information and difficult interpretability, and therefore from these raw data signals, we performed data and domain-driven feature engineering to generate potential predictor variables for our model that were easily interpretable and well represented the information in the data. We generated a total of 50 features across all the raw signal variables.

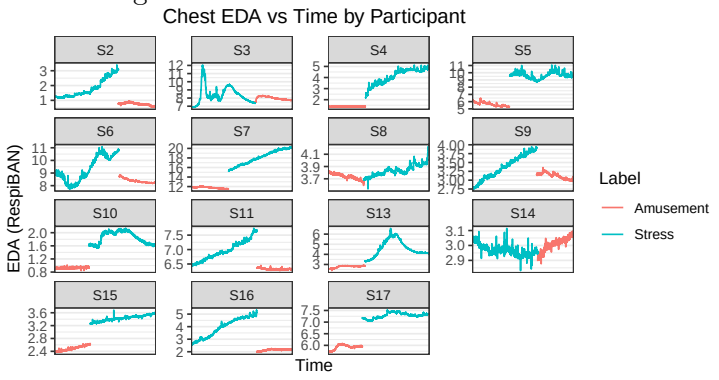

Fig 1.1 Plot of RespiBAN (Chest) EDA samples collected over time by participant and label (2–stress, 3–amusement)

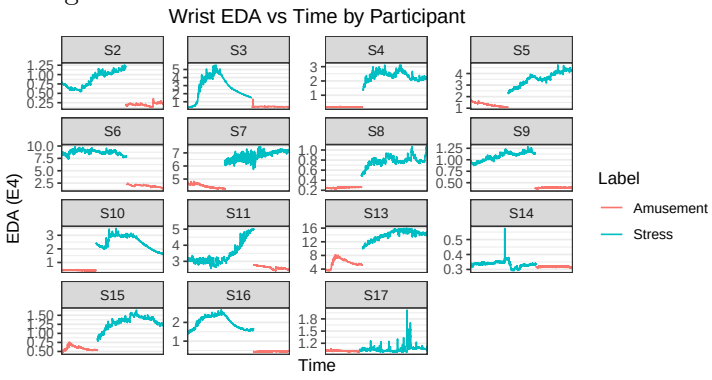

Fig 1.2 Plot of E4 (Wrist) EDA samples collected over time by participant and label (2–stress, 3–amusement)

4 Methodology

4.1 Feature Engineering

As mentioned earlier, the brain takes .9 milliseconds to shift the body into a full stress response[4], which suggests that these data should be sampled at approximately 1111 Hz to properly identify a state of stress. However, the highest-resolution data is sampled at 700 Hz, which suggests that these data are not granular enough to fully capture a shift to stress. Furthermore, we extracted features at a higher precision but found no significant differences than ones extracted at 4Hz. As our study is focused on determining the stress detection over long periods of time and with computational cost in mind, we decided to downsample our data to 4Hz. We assigned a unique ID to each of the 15 subjects, and also added their personal characteristics. Since we are primarily interested in differentiating between stress and amusement, we only kept samples that were labeled 2 (stress) or 3 (amusement). We then relabeled amusement as 0 and stress as 1.

Next, we calculated features for the different modalities. Segmentation of the physiological signals was done using a sliding window, with a window shift of 0.25 seconds. The ACC and physiological features were all extracted using a 5 second window size [7][8]. To calculate some BVP and ECG features, we utilized peak detection algorithms [9] to identify the time between each successive peaks and used the peak intervals to calculate the mean, standard deviation and root-mean squared of heart rate variability within each assigned windows.

To calculate the respiration variables, we had to separate out inhales and exhales. Inhale data corresponded to respiration values greater than zero and exhale data corresponded to respiration values less than zero. For five-second intervals where someone either did not inhale or exhale, we coded the corresponding observations to equal zero instead of NA. Additionally, we found total respiration range which took both inhale and exhale values

into account when calculating the maximums and minimums. Next, we calculated the inhale to exhale ratio by dividing our mean inhale column by our mean exhale column. Since this would have produced some infinite values due to the fact that periods where someone was not exhaling were set to zero, we set these inhale/exhale ratios to equal 10. Additionally, we had a couple of intervals where the inhale/exhale ratio was very large, sometimes even greater than 1500. This was happening because some five-second windows in our data could have consisted of the participant inhaling for most of the time, thus we would be dividing by an extremely low exhale value. So, we re-coded any ratio greater than 10 to equal an upper bound of 10. This is a fair assumption, since it would not be physically possible for someone to inhale incredibly more than they exhale. We next calculated a breath rate by first summing the number of rows in one breath cycle, which is defined as one full inhale and one full exhale. Then, we took this sum and divided it by four, since the observations in our dataset were taken every quarter of a second, to get the time in seconds of one full breath cycle.

For ACC, we also calculated the 3D magnitude, $ACC = \sqrt{ACC_X^2 + ACC_Y^2 + ACC_Z^2}$. We then calculated the mean, standard deviation, maximum, minimum and range of these modalities; a summary of all predictor variables is given in Table 1.

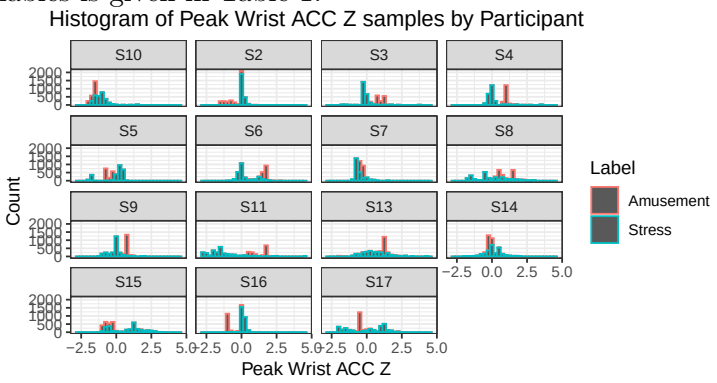

Fig 1.3 Histogram of peak acc wrist samples by participant and label (2−stress, 3−amusement)

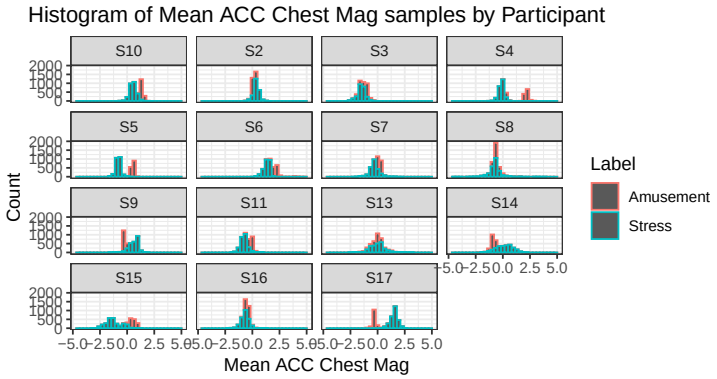

Fig 1.4 Histogram of mean acc chest magnitude samples by participant and label (2−stress, 3−amusement)

Performing exploratory data analysis on the engineered features provided more insightful information related to our study. For example, Figure 1.3 and 1.4 are histograms of Peak Wrist ACC Z and Mean ACC Chest Magnitude samples engineering from the raw signal ACC data. From these plots we can observe that these ACC feature values have different distributions based on the different conditions and could potentially be significant in predicting stress or amusement.

Other interesting insights like potential issues of multicollinearity were observed when plotting correlation matrices of the features. Figure 2 shows features engineered only from EDA and we can see that there exists a high level of collinearity. It is important to note that some correlation values are 1.00 due to the large size of our dataset and having the features engineered from the same signal variable. On the other hand, there was low correlation between features of different signal variables. High collinearity in predictor variables will influence the strategies and thresholds chosen for model selection and interaction effects, which we elaborate further in the paper in relation to our proposed model.

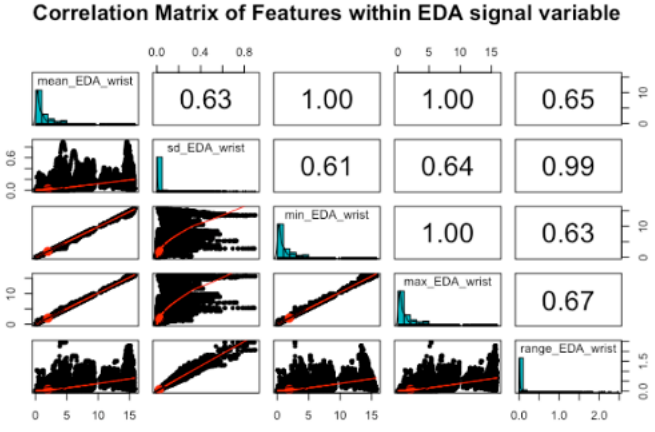

Fig 5.3 Correlation Matrix of Features within EDA signal variable

Table 1: Summary of Sensor Data Categories, with the Extracted Predictors and Respective Symbols

Metric	Features
Chest Sensor Data	
Chest ACC	Mean (μ^C_{ACC}), SD (σ^C_{ACC})
Chest ECG	Mean (μ^C_{ECG}), SD (σ^C_{ECG}), Mean HRV (μ^C_{HRV}), SD HRV (σ^C_{HRV})
Chest EDA	Mean (μ^C_{EDA}), SD (σ^C_{EDA}), Min (Min^C_{EDA}) Max (Max^C_{EDA}), Range ($Range^C_{EDA}$)
Chest EMG	Mean (μ^C_{EMG}), SD (σ^C_{EMG}), Min (Min^C_{EMG}) Max (Max^C_{EMG}), Range ($Range^C_{EMG}$)
Chest Resp	Mean (μ^C_{inhale}), SD (σ^C_{inhale}) Mean (μ^C_{exhale}), SD (σ^C_{exhale}) Min (Min^C_{resp}), Max (Max^C_{resp}), Range ($Range^C_{resp}$) Mean (μ^C_{breath}), SD (σ^C_{breath}), Min (Min^C_{breath}) Max (Max^C_{breath}), Range ($Range^C_{breath}$), I/E Ratio
Chest Temp	Mean (μ^C_T), SD (σ^C_T), Min (Min^C_T), Max (Max^C_T), Range ($Range^C_T$)
Wrist Sensor Data	
Wrist ACC	Mean (μ^W_{ACC}), SD (σ^W_{ACC})
Wrist BVP	Mean (μ^W_{ECG}), SD (σ^W_{ECG}), Mean HRV (μ^W_{HRV}), SD HRV (σ^W_{HRV})
Wrist EDA	Mean (μ^W_{EDA}), SD (σ^W_{EDA}), Min (Min^W_{EDA}), Max (Max^W_{EDA}), Range ($Range^C_{EDA}$)
Wrist Temp	Mean (μ^W_T), SD (σ^W_T), Min (Min^W_T), Max (Max^W_T), Range ($Range^W_T$)

4.2 Model Selection

We first split our data into a 80/20 train/test split, perform 5-fold cross validation on the training data and lastly assess model performance by predicting on the testing data.The extracted features serve as predictors for our classification models. We propose two different models, both using different types of sensor data: features from the wrist-worn device, and features from both chest-worn and wrist-worn devices. Since our research goals are concerned with the predictive power and interpretability of our models, logistic regression was chosen over other classification algorithms. To determine our final model, we performed variable selection by first including all relevant features and *participant* and then used backward selection with AIC as the selection criterion. We also tried forward selection and the BIC criterion, but ended up with similar models. Then, we iteratively removed variables that had p-values greater than 0.1. Interaction effects between the remaining predictors and with *participant* were fit to determine if they would improve the predictive power of the models. We also noted that the models with the subjects' personal characteristics did not increase accuracy, F_1-score or AUC values, and also led to increased computational cost. Thus, we did not include any personal characteristics in the final models.

We also tried fitted multilevel logistic regression models, where *participant* was the random effect. However, these models were computationally expensive, and did not improve the classification metrics. We also utilized the random forest algorithm; these models resulted in better predictions, but lack sufficient interpretation. Both of these algorithms are discussed in the Appendix.

The final models are below (Model 1 was fitted using only the wrist-worn data and Model 2 was fitted using all the sensor data). i corresponds to each sample, and Y_i indicates the corresponding state (1 = stress or 0 = amusement). $subject_i$ represents the unique subject ID that was assigned to each of the 15 participants. The reference level for *subject* is Subject 10 (S10), who had the highest number of samples in the filtered dataset. The remaining levels $j \in (2, 3, \ldots, 14, 15)$ correspond to S2, S3, S4, S5, S6, S7, S8, S9, S11, S13, S14, S15, S16 and S17 respectively. The other predictors correspond to those displayed in Table 1.

$$Y_i \sim Bernoulli(\pi_i)$$

$$log(\frac{\pi_i}{1-\pi_i}) = \beta_0 + \beta_1\ \sigma^W_{iEDA} + \beta_2\ \sigma^W_{iACC} + \beta_3\ \sigma^W_{iHRV}\ + \beta_4\ \sigma^W_{iEDA} * \sigma^W_{iHRV}\ + \beta_5\ \sigma^W_{iEDA} * \sigma^W_{iACC}\ +$$
$$\beta_6\ \sigma^W_{iHRV} * \sigma^W_{iACC}\ + \sum_{j=2}^{15} \beta_{7j}\ I(subject_i = j)\ + \sum_{j=2}^{15} \beta_{8j}\ \sigma^W_{iACC} * I(subject_i = j) \quad (1)$$

$$Y_i \sim Bernoulli(\pi_i)$$

$$log(\frac{\pi_i}{1-\pi_i}) = \beta_0 + \beta_1\ \mu^C_{iACC} + \beta_2\ \sigma^C_{iACC} + \beta_3\ \sigma^W_{iACC} + \beta_4\ \sigma^C_{iHR} + \beta_5\ \sigma^W_{iHR}\ + \beta_6\ \sigma^W_{iHRV}\ +$$
$$\beta_7\ rms^C_{iHRV} + \beta_8\ rms^W_{iHRV} + \beta_9\ \sigma^C_{iEDA}\ + \beta_{10}\ \sigma^W_{iEDA} + \beta_{11}\ min^C_{iEMG} + \beta_{12}\ \sigma_{i\ inhale}\ +$$
$$\beta_{13}\ \sigma_{i\ exhale} + \beta_{i14}\ max_{i\ breath} + \beta_{15}\ ie_{i\ ratio} + \beta_{16}\ \sigma^C_{iT} + \beta_{17}\ min^W_{iT} + \beta_{18}\ \sigma^C_{iEDA} * min^W_{iT}\ +$$
$$\sum_{j=2}^{15} \beta_{19j}\ I(subject_i = j) + \sum_{j=2}^{15} \beta_{20j}\ rms^W_{iHRV} * I(subject_i = j)\ + \sum_{j=2}^{15} \beta_{21j}\ \sigma^C_{iHR} * I(subject_i = j) \quad (2)$$

4.3 Evaluation Metrics

We used accuracy, F_1-score, and AUC as evaluation metrics. Accuracy is the number of correctly classified observations over the total number of samples. We then define the F_1-score and AUC. The True Positive Rate (TPR/Recall) is the proportion of positive (stress) samples that are correctly identified. The False Positive Rate (FPR) is the proportion of negative (amusement) samples that are predicted to be positive over the total number of negative samples. Precision is the proportion of samples that are positive and classified as one over the total number of samples that are predicted to be positive. Thus, the F_1-score is the harmonic average of the precision and recall, and is a good choice for imbalanced classification tasks. The Receiver Operating Characteristics (ROC) curve plots the TPR against the FPR. The AUC is thus the area under the ROC curve. Both the F_1-score and AUC have a range of [0,1], where a higher value indicates better classification, and a value of 1 indicates perfect classification.

5 Results

Table 2: Model Results with 5-fold Cross Validation, with Mean and Standard Deviation reported

	Cross Validation Data			Test Data		
Model	Accuracy	F1-Score	AUC	Accuracy	F1-Score	AUC
Wrist Only	85.83% ± 0.45%	79.81 ± 0.77	84.17 ± 0.64	86.2%	80.45	84.70
Chest & Wrist	90.88% ± 0.05%	87.23 ± 0.08	90.05 ± 0.13	90.83%	87.27	90.19

5.1 Wrist Sensor Data

Using only data from the wrist-worn wearable, our model reports an accuracy of 86.2%, an F_1-score of 80.45 and an AUC of 84.7 when predicting on the testing data. This suggests that we can detect stress vs. amusement reasonably well using only the wrist-worn wearable, which may be preferable since a wrist-worn device is minimally intrusive and more convenient than a chest-worn device. A naive classifier using only wrist sensor data that predicts only the majority class from each test fold reports an accuracy of 64.26 ± 0.00; this indicates our model exceeds that of a naive classifier.

5.2 All Sensor Data

Using data from both the chest-worn and wrist-worn wearables, our model reports an accuracy of 90.83%, an F_1-score of 87.27 and an AUC of 90.19 when predicting on the testing data. These values suggest that sensor data are useful in discriminating between stress and amusement conditions. These metrics are slightly better than the previous model, which indicates that having data from both the chest-worn and wrist-worn wearables is better than having just data from the wrist-worn device. A naive classifier using all sensor data that predicts only the majority class from each test fold reports an accuracy of 64.26 ± 0.00; similarly, this indicates our model exceeds that of a naive classifier.

5.3 Model Diagnostics

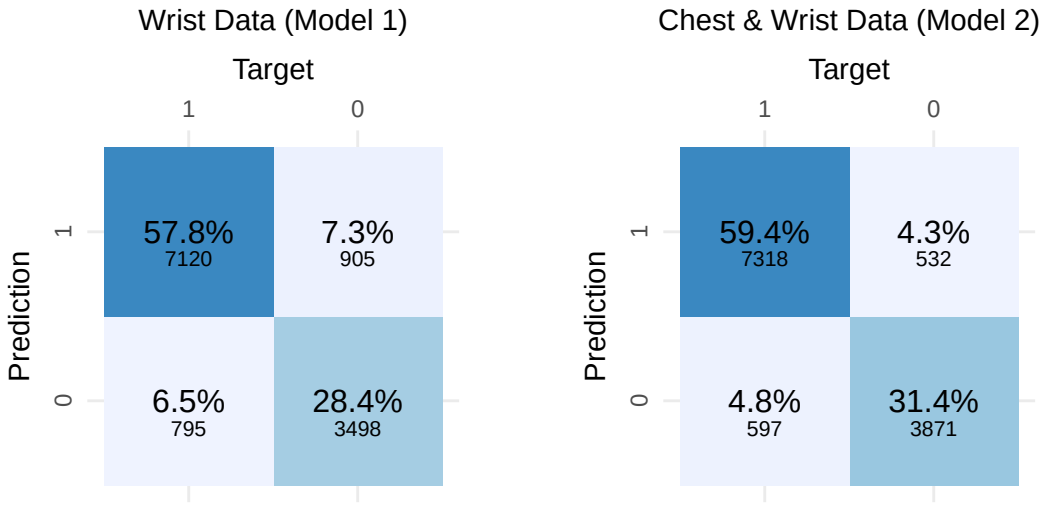

Figure 3: Confusion Matrices on Model Predictive Accuracy for model trained on only Wrist Data and Wrist and Chest data combined

From the metrics displayed in Table 2 and the confusion matrices in Figure 3, it seems that both the wrist model and chest & wrist models have relatively high classification accuracies when detecting stress vs. amusement.

It is also worth discussing independence when evaluating our logistic regression model. Adding *participant* addresses independence at the subject level, but each individual sample is not independent; presumably, a person's heart rate from one second ago affects their current heart rate. In addition, we display variance inflation factor (VIF) values for both our models in the Appendix in Tables 9-11. VIF quantifies the severity of multicollinearity in a logistic regression model, and a smaller VIF value indicates less multicollinearity. In general, most VIF values are below 10; however, *participant* and predictors that are used in interaction effects may have higher VIF values. However, since the standard errors of our estimates and interaction effects are relatively small, the larger VIF values are acceptable.

5.3.1 Sensitivity Analysis

Table 3: Model Results with Data Sampled using 60s window size and 0.25s window shift

	Cross Validation Data			Test Data		
Model	Accuracy	F1-Score	AUC	Accuracy	F1-Score	AUC
Wrist Only (Model 3)	87.59% ± 0.34%	82.9 ± 0.44	86.97 ± 0.36	87.3%	82.46	86.60
Chest & Wrist (Model 4)	93.32% ± 0.3%	90.58 ± 0.43	92.69 ± 0.35	93.03%	90.13	92.29

Table 3 displays the results of the wrist data and all data models (Models 3 and 4), with the same predictors as Models 1 and 2, when using data that was sampled using a 60 second window size and a 0.25 second window shift (as compared to a 5 second window size and 0.25 second window shift). In general, this set of models performed better in terms of evaluation metrics. Similarly, we compare the estimates from this set of models to the estimates

from our original models; these estimates are displayed in Tables 7 and 8 in the Appendix. Our wrist-only model appears to be reasonably robust, as most estimates using data sampled using a 60 second window size and a 0.25 second window shift were similar to the estimates from Model 1. However, the combined data model was less robust, as there are estimates, such as the one for the standard deviations of chest EDA that were quite different (1.12 vs. 2.36 for Model 2 vs. Model 4). This disparity could be due to the loss of data when sampling using a 60 second window, the choice to initially downsample our data to 4Hz or other external factors that these features could not capture in these models.

5.3.2 Comparison to Other Models

In addition to our logistic regressions, we also ran a random forest model on the data. In our wrist-only random forest model, we found that min wrist EDA, max wrist EDA, range of wrist EDA and standard deviation of Z wrist ACC were the only significant predictors. While we don't see these exact predictors in our logistic regression, we had standard deviation of wrist EDA and standard deviation of wrist ACC in our final model, and these variables are highly correlated with the ones in our random forest model. For the model built with chest and wrist data, we saw that max wrist EDA and standard deviation of chest ACC were the only two important predictors outside of participant. Standard deviation of chest ACC also occurs in our logistic regression, however instead of max wrist EDA, we see standard deviation of wrist EDA. Again, these two variables are highly correlated, so that seems to make sense. Further discussion of our random forests models can be found in the appendix.

6 Discussion

6.1 Interpretation of Wrist-Only Model

When standard deviation of wrist EDA increases by 1, we expect the odds of being stressed versus amused to multiply by a factor of 45.879 (with 95% CI 39.997, 52.625), holding all else constant. When standard deviation of wrist ACC increases by 1, we expect the odds of being stressed versus amused to multiply by a factor of 1.56 (15.812, 99.409), holding all else constant. When standard deviation of wrist HRV increases by 1, we expect the odds of being stressed versus amused to multiply by a factor of 56.092 (1.429, 1.704), holding all else constant.

Holding standard deviation of wrist HRV constant, when standard deviation of wrist EDA and wrist ACC both increase by 1, we expect the odds of being stressed versus amused to multiply by an additional factor of 0.323 (0.304, 0.344). Holding standard deviation of wrist ACC constant, when standard deviation of wrist EDA and wrist HRV both increase by 1, we expect the odds of being stressed versus amused to multiply by an additional factor of 1.649 (1.401, 1.94). Holding standard deviation of wrist EDA constant, when standard deviation of wrist HRV and wrist ACC both increase by 1, we expect the odds of being stressed versus amused to multiply by an additional factor of 1.288 (1.118, 1.483).

6.2 Interpretation of Chest and Wrist Model

When standard deviation of chest EDA increases by 1, we expect the odds of being stressed versus amused to multiply by a factor of 3.313 (2.957, 3.712), holding all else constant. When standard deviation of wrist EDA increases by 1, we expect the odds of being stressed versus amused to multiply by a factor of 6496.378 (4860.614, 8682.632), holding all else constant. When standard deviation of chest temperature increases by 1, we expect the odds of being stressed versus amused to multiply by a factor of 1.186 (1.143, 1.232), holding all else constant. When minimum wrist temperature increases by 1, we expect the odds of being stressed versus amused to multiply by a factor of 0.165 (0.15, 0.182), holding all else constant. When minimum chest EMG increases by 1, we expect the odds of being stressed versus amused to multiply by a factor of 0.867 (0.832, 0.903), holding all else constant. When standard deviation of inhale increases by 1, we expect the odds of being stressed versus amused to multiply by a factor of 1.774 (1.685, 1.866), holding all else constant. When standard deviation of exhale increases by 1, we expect the odds of being stressed versus amused to multiply by a factor of 0.766 (0.736, 0.799), holding all else constant. When maximum breath cycle time increases by 1, we expect the odds of being stressed versus amused to multiply by a factor of 1.953 (1.889, 1.827), holding all else constant. When inhale/exhale ratio increases by 1, we expect the odds of being stressed versus amused to multiply by a factor of 0.749 (0.722, 0.777), holding all else constant. When mean chest ACC increases by 1, we expect the odds of being stressed versus amused to multiply

by a factor of 1.04 (0.99, 1.092), holding all else constant. When standard deviation of chest ACC increases by 1, we expect the odds of being stressed versus amused to multiply by a factor of 2.23 (2.111, 2.356), holding all else constant. When standard deviation of wrist ACC increases by 1, we expect the odds of being stressed versus amused to multiply by a factor of 1.474 (1.374, 1.582), holding all else constant. When standard deviation of wrist HR increases by 1, we expect the odds of being stressed versus amused to multiply by a factor of 1.232 (1.176, 1.292), holding all else constant. When standard deviation of wrist HRV increases by 1, we expect the odds of being stressed versus amused to multiply by a factor of 1.189 (1.143, 1.236), holding all else constant. When root mean square of wrist HRV increases by 1, we expect the odds of being stressed versus amused to multiply by a factor of 1.326 (1.186, 1.482), holding all else constant. When standard deviation of chest HR increases by 1, we expect the odds of being stressed versus amused to multiply by a factor of 0.477 (0.396, 0.575), holding all else constant. When standard deviation of chest HR increases by 1, we expect the odds of being stressed versus amused to multiply by a factor of 1.063 (1.022, 1.105), holding all else constant.

Additionally, holding all else constant, when standard deviation of wrist HRV and minimum wrist temperature both increase by 1, we expect the odds of being stressed versus amused multiply by an additional factor of 0.009 (0.008, 0.011).

It is also interesting to note that in the model built with chest and wrist data, most of our predictors appear to be standard deviation measures, and in the wrist-only model, participant is the only non-standard deviation predictor. This observation suggests that the variability in physiological signals is a more important distinguisher of stress and amusement as opposed to the mean or maximum of these signals in many cases.

6.3 Quantifying Heterogeneity

Please see Figures 3.1-3.5 in the Appendix for a graphical view of the Confidence Intervals for participant and participant interaction effects relative to our baseline of Subject 10. These estimates are graphed in terms of the log-odds for scaling reasons, however, you can see all of these intervals in terms of odds in Tables 7 and 8 in the Appendix.

For the wrist only model, we see that Subjects 2, 3, 4, 6, 8, 11, 15, 16, and 17 are all less likely to be stressed if physiological measures are at the mean for the dataset than Subject 10. However, all of these intervals overlap except for those of Subjects 6 and 13, indicating that the other subjects don't differ significantly from each other. Additionally, Subject 14's log-odds of being stressed do not differ from Subject 10, and Subjects 5, 7, and 9 all appear to have greater odds of being stressed than Participant 10, and these three people differ significantly from each other.

The interaction effect between participant and standard deviation of ACC appears to be significantly higher for Subjects 5 and 7 compared to the baseline of Subject 10. Subjects 9 and 14's high positive estimates also indicate that as standard deviation of ACC increases, their log-odds of being stressed increase at a greater rate than we see for the baseline Subject 10, however these estimates appear not to be significant. For the remaining subjects (2, 3, 4, 6, 8, 11, 13, 15, 16, and 17), their log-odds of being stressed decrease at a greater rate as standard deviation of ACC increases compared to Subject 10, however their confidence intervals all overlap, so they don't appear to differ significantly from each other.

In the chest and wrist model, Subjects 17 and 9 have significantly greater log-odds of being stressed compared to Subject 10 when all other variables are at their mean. Subjects 2, 14, and 15 do not differ significantly from Subject 10. Subjects 3, 4, 5, 6, 7, 8, 11, 13, and 16 all have significantly lower log-odds of being stressed compared to Subject 10, but they don't all differ significantly from each other. Subjects 6 and 13 have the lowest log-odds, followed by Subjects 7 and 11, Subjects 3, 4, and 5, and finally subjects 8 and 16.

For the interaction effect between participant and root mean square wrist HRV, as root mean square HRV increases, Subject 11's log-odds of being stressed increase more compared to the baseline Subject 10, and Subject 3's log-odds also significantly increase more, although not by as much as Subject 11. Subjects 4 and 16 do not appear to have a significant difference in the impact on stress as root mean square HRV increases compared to Subject 10. We expect Subjects 2, 5, 6, 8, 9, and 14 all to have a decrease in log-odds of being stressed as root mean square HRV increases compared to Subject 10, however these subjects do not seem to vary significantly between each other. Lastly, Subjects 7, 13, 15, and 17 all see an even greater decrease in log-odds of being stressed as root mean square HRV increases compared 2, 5, 6, 8, 9, and 14, but again, and Subject 7 varies significantly from Subjects 15 and 17, but not 13.

Lastly, for the interaction effect between participant and standard deviation of chest HR, Subject 8 is the only subject whose log-odds of being stressed decrease as sd HR increases, relative to Subject 10. Subjects 2 and 5 do not have significantly differing effects of sd chest HR on the log-odds of being stressed compared to Subject 10. The remaining Subjects (9, 7, 6, 4, 3, 17, 16, 15, 14, 13, and 11) all have increasing log-odds of being stressed as sd HR increases compared to Subject 10, however not all of them significantly differ from each other. Subject 11 is significantly higher than all of the other subjects, and Subject 17 significantly differs from Subjects 3, 7, 9, 14, 15, and 16.

We see especially large coefficients for Subjects 5 and 7 in the wrist model, and this is possibly due to perfect seperability. This means that the physiological measures have little or no overlap between the stress and amusement conditions, and since this is the case, it can give us some stranger-looking model estimates. Collecting more data or measuring stress vs. amusement in the real world instead of a laboratory study might be ways to tackle this issue in the future.

7 Conclusion

While we were able to distinguish between stress and amusement conditions fairly well with our model, it is worth pointing out some limitations. First, we only had data for the 15 participants that have similar ages and are mostly male. As participant ID is an important predictor in our models, we are unable to generalize the likelihood of being stressed or amused to other age groups or identified genders. Also, since all of the data was collected in one sitting during an experiment, our data is very highly correlated and suggests there is heterogeneity across individuals' responses to stress vs. amusement. We can distinguish between stress and amusement in this controlled laboratory setting, but we are unable to generalize the differences in these emotions in a person's day-to-day life, or what stress and amusement might even look like if the measurements were taken on a different day.

Since it appears that the wrist wearable is sufficient for detecting stress versus amusement in the laboratory setting, further research could study wrist data for individuals over a long time period, perhaps a day or even a week, and see if different emotions are still distinguishable. Also, studies could be done with a larger random sample of individuals to try and see if there are any findings about stressed versus amused physiological responses generalizable to all individuals instead of just at the participant level.

8 Appendix

8.1 Bibliography

1. D. Watson, L. A. Clark and A. Tellegen. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. Journal of Personality and Social Psychology, 54, 6, 1063.
2. P. Schmidt et al. (2018). Introducing WESAD, a Multimodal Dataset for Wearable Stress and Affect Detection. Proceedings of the 2018 on International Conference on Multimodal Interaction - ICMI 18.
3. H. Yaribeygi et al. (2017). The impact of stress on body function: A review. EXCLI Journal, 2017, 16, pp. 1057–1072.
4. D. Strickland. (2019). .9 Milliseconds. Transitions.
5. L. Shu et al. (2018). A Review of Emotion Recognition Using Physiological Signals. Sensors (Basel), 18, 7.
6. D. Girardi et al. (2017). Emotion Detection Using Noninvasive Low Cost Sensors
7. M. Arif and A. Kattan. (2015). Physical Activities Monitoring Using Wearable Acceleration Sensors Attached to the Body. PLoS ONE 10, 7.
8. G. Uy et al. (2013). Correlation of Stress Inducers and Physiological Signals with Continuous Stress Annotations. WCTP 2012, PICT 7, pp. 178–183.
9. P.Westermark (2015). Package 'peakPick'. https://cran.r-project.org/web/packages/peakPick/peakPick.pdf

8.2 Random Forest Model

In addition to the logistic regression, we also fit several other types of models. The first of those was a random forest model. We built a random forest model using all of our data and just the wrist data, and had great overall performance for both. Our model built with all the data consisted of only three predictors: participant ID, max wrist electrodermal activity, and magnitude of standard deviation of chest movement, and it had a misclassification rate of 0.74%. The model built using only wrist data used five predictors, which included participant ID, minimum electrodermal activity, maximum electrodermal activity, the range of electrodermal activity, and standard deviation of Z-plane movement, and it had an even lower misclassification rate of 0.002%.

Despite the great performance of these models, we opted to go with the logistic regression as our final model since interpretability was important to the goals of the case study. Random forest is great for prediction, however it is difficult to interpret how exactly the predictors affect the response. Therefore, we were willing to sacrifice a little bit of accuracy in order to make more concrete conclusions about how our predictor variables affected stress and amusement.

8.3 Multilevel Model

Another method we tried was using a multilevel model (MLM) with the features engineered from the raw signal data being the random effects of the model, while survey and participant-specific data as the fixed effects. One large issue with running MLM was the computational intensity it took after including more than 2 fixed effects and 4 random effects. Considering the number of predictor variables we wished to include and analyze, we decided that it was not feasible to run a multi-level model due to the nature of our study and dataset having a large number of samples and predictor variables.

8.4 Heterogeneity Plots

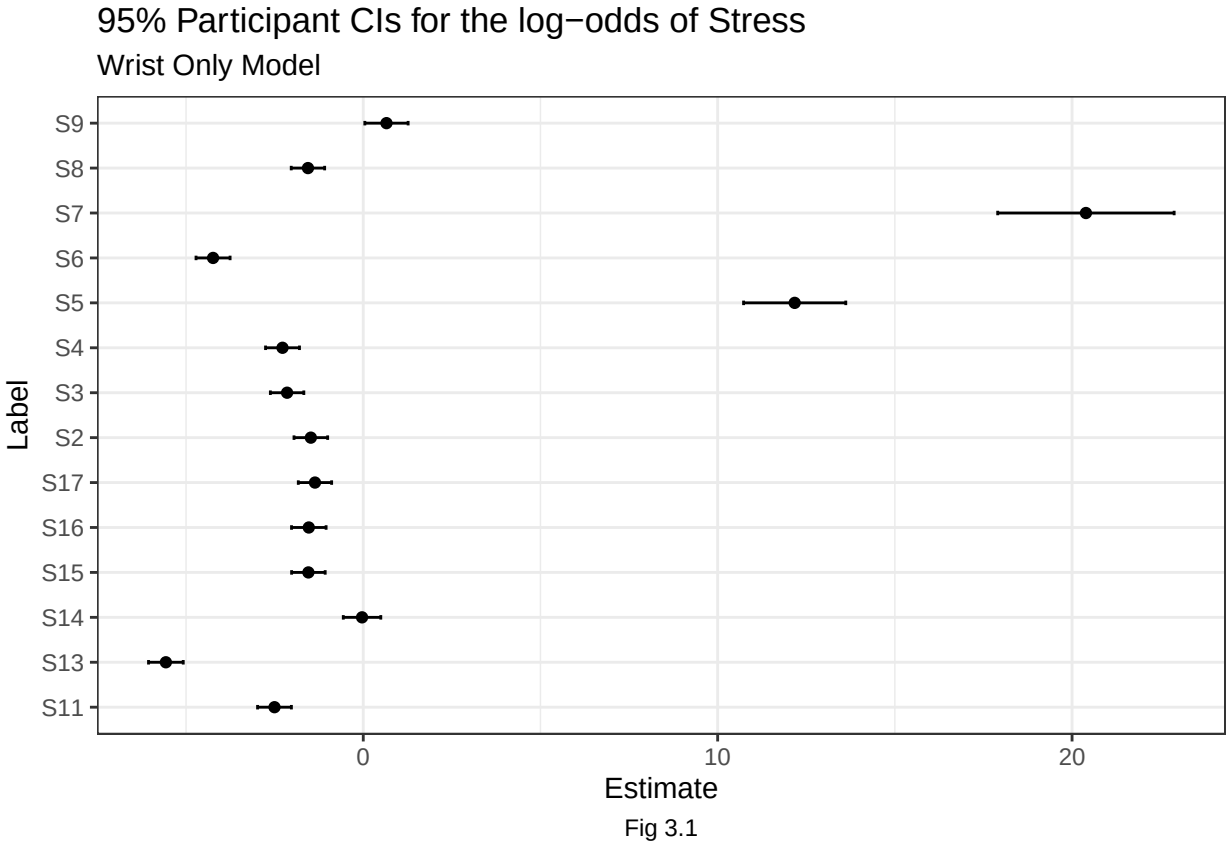

Fig 3.1

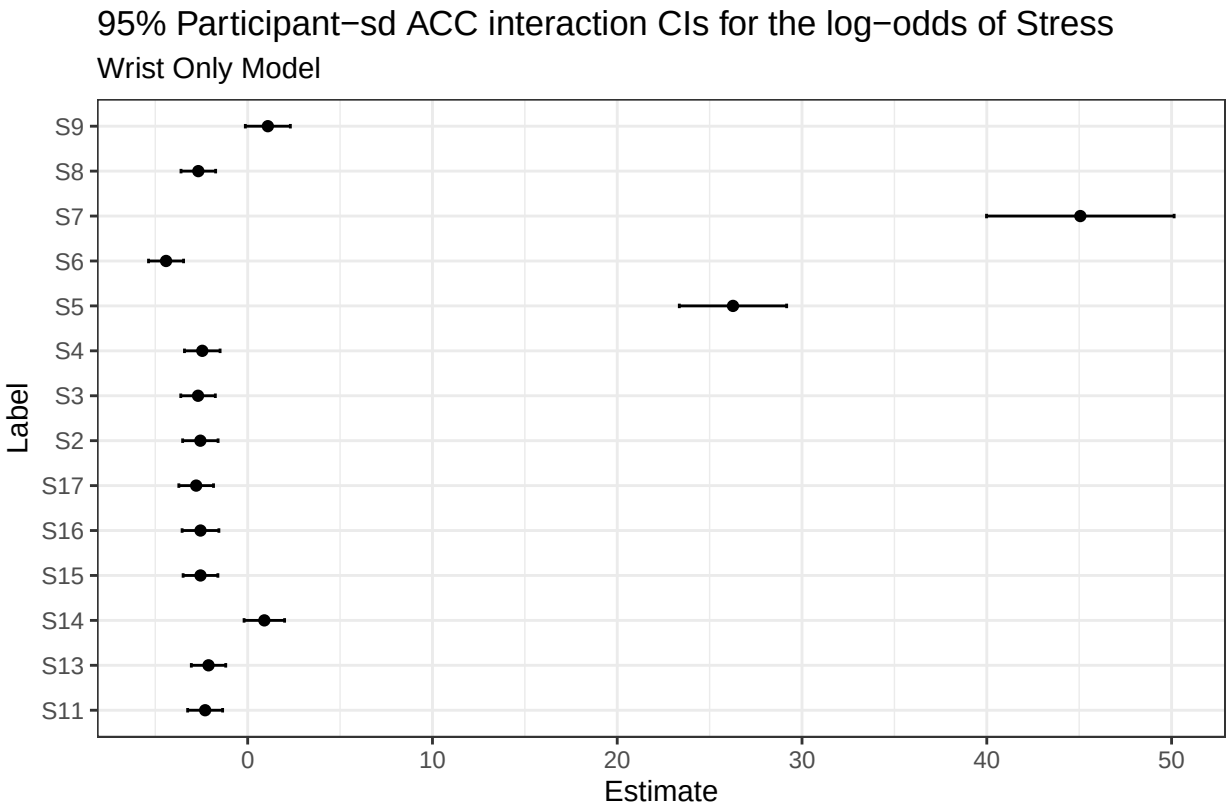
95% Participant−sd ACC interaction CIs for the log−odds of Stress
Wrist Only Model
Label
S9
S8
S7
S6
S5
S4
S3
S2
S17
S16
S15
S14
S13
S11
0
10
20
30
40
50
Estimate

Fig 3.2

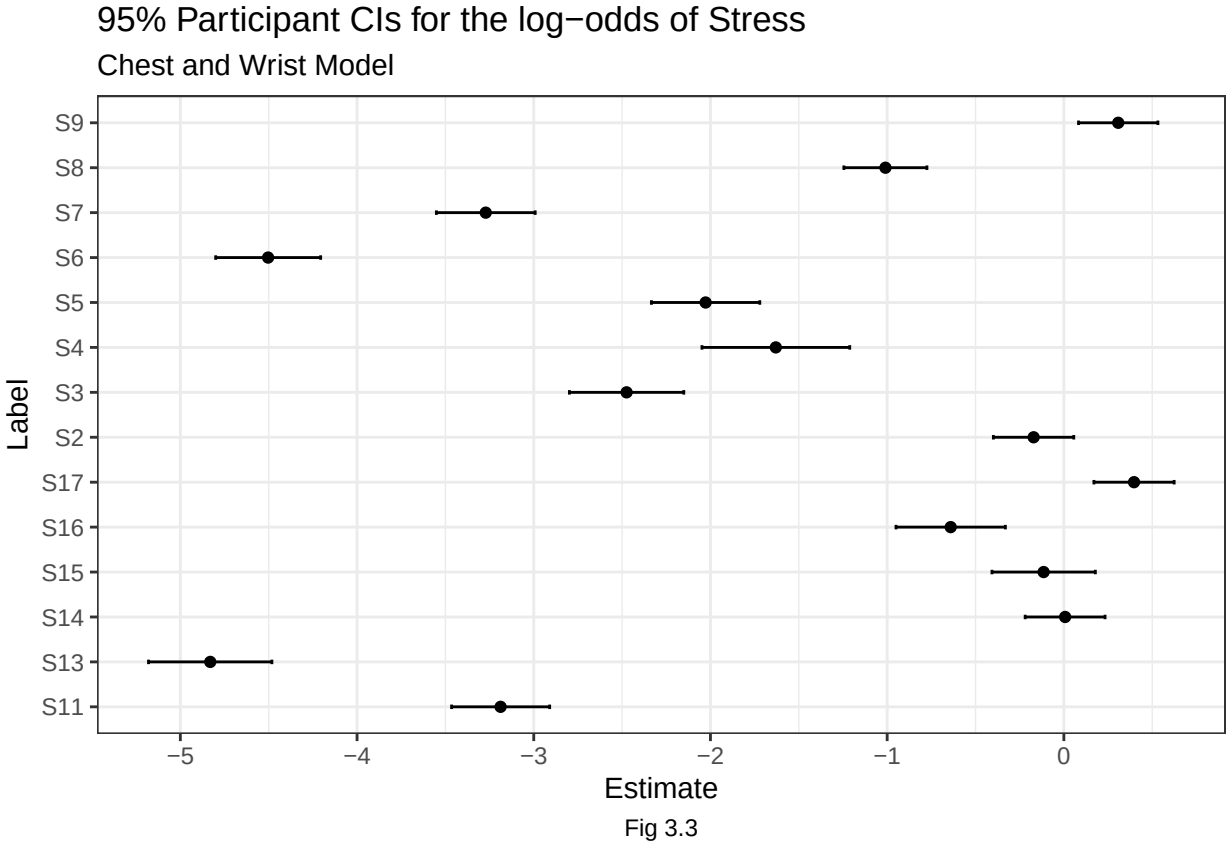
95% Participant CIs for the log−odds of Stress
Chest and Wrist Model
Label
S9
S8
S7
S6
S5
S4
S3
S2
S17
S16
S15
S14
S13
S11
−5
−4
−3
−2
−1
0
Estimate

Fig 3.3

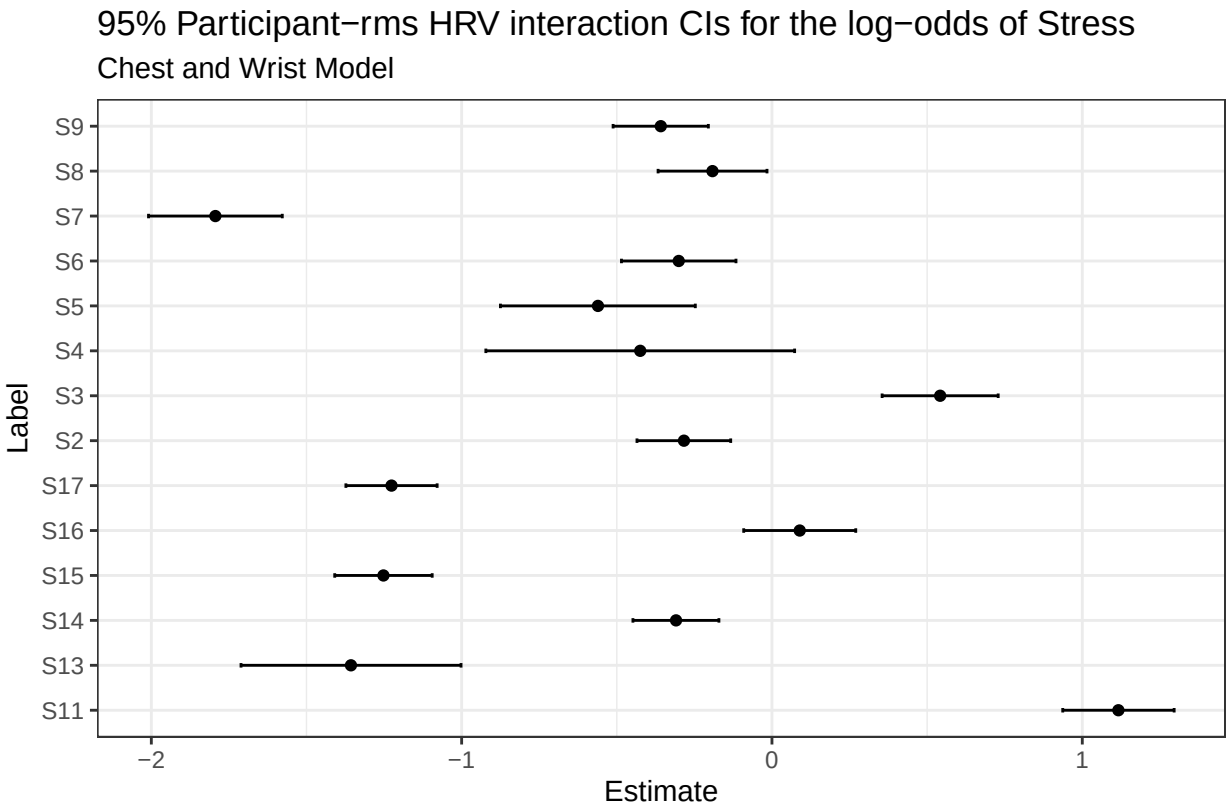

Fig 3.4

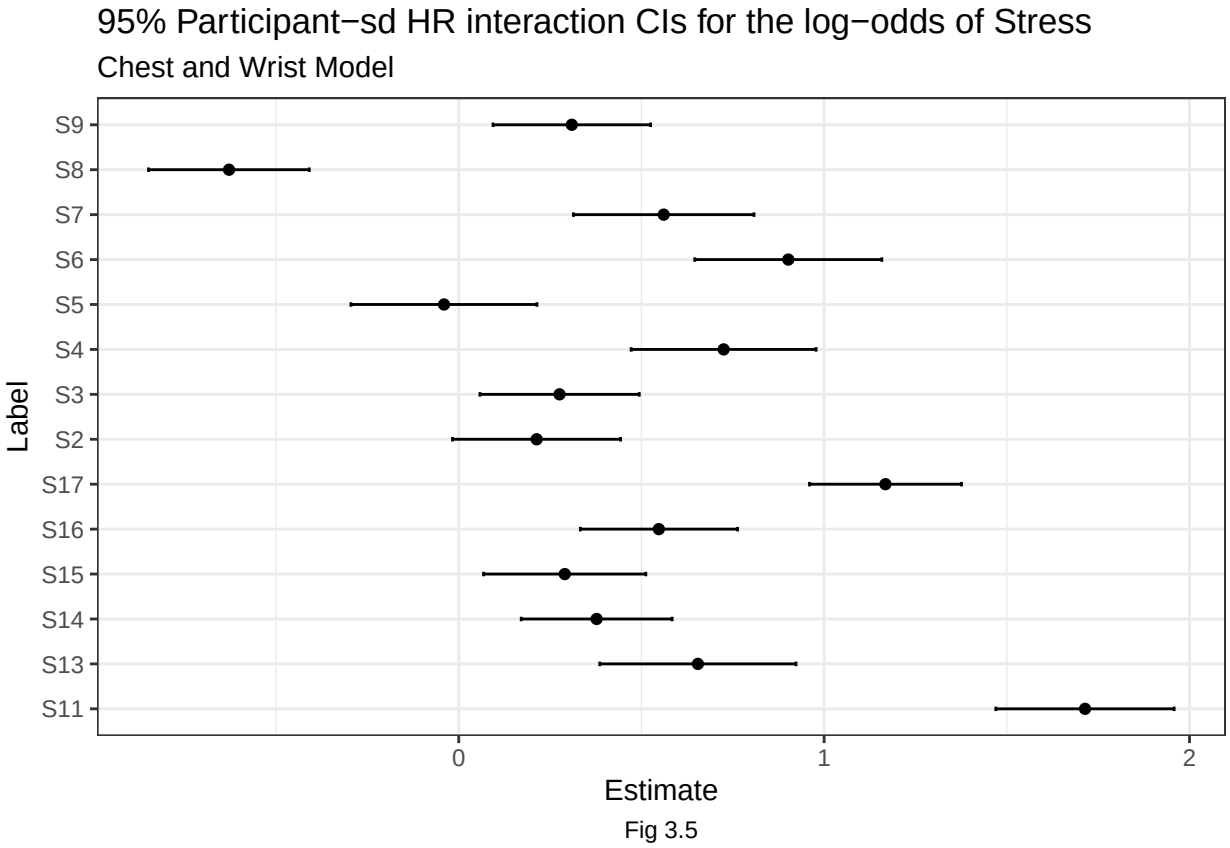

Fig 3.5

8.5 Model Output

8.6 VIF

Table 4: Coefficients obtained with Model 1 (Wrist Model, Data Sampled at 4Hz)

Coefficient	Estimate	Std. Error	p-value	Lower CI	Upper CI
(Intercept)	4.027	0.236	< 2e-16	3.564	4.49
σ^W_{EDA}	3.826	0.07	< 2e-16	3.689	3.963
σ^W_{ACC}	3.68	0.469	< 2e-16	2.761	4.599
σ^W_{HRV}	0.445	0.045	< 2e-16	0.357	0.533
σ^W_{EDA} * σ^W_{ACC}	-1.129	0.031	< 2e-16	-1.19	-1.068
σ^W_{EDA} * σ^W_{HRV}	0.5	0.083	< 2e-16	0.337	0.663
σ^W_{ACC} * σ^W_{HRV}	0.253	0.072	< 2e-16	0.112	0.394
Participant					
S10	(Reference)				
S2	-1.479	0.242	< 2e-16	-1.953	-1.005
S3	-2.149	0.241	< 2e-16	-2.621	-1.677
S4	-2.279	0.244	< 2e-16	-2.757	-1.801
S5	12.174	0.736	< 2e-16	10.731	13.617
S6	-4.238	0.245	< 2e-16	-4.718	-3.758
S7	20.392	1.27	< 2e-16	17.903	22.881
S8	-1.56	0.241	< 2e-16	-2.032	-1.088
S9	0.656	0.312	0.035	0.044	1.268
S11	-2.505	0.243	< 2e-16	-2.981	-2.029
S13	-5.57	0.25	< 2e-16	-6.06	-5.08
S14	-0.034	0.27	0.899	-0.563	0.495
S15	-1.548	0.242	< 2e-16	-2.022	-1.074
S16	-1.537	0.249	< 2e-16	-2.025	-1.049
S17	-1.362	0.24	< 2e-16	-1.832	-0.892
σ^W_{ACC} * **Participant**					
σ^W_{ACC} * S10	(Reference)				
σ^W_{ACC} * S2	-2.561	0.489	< 2e-16	-3.519	-1.603
σ^W_{ACC} * S3	-2.686	0.477	< 2e-16	-3.621	-1.751
σ^W_{ACC} * S4	-2.457	0.491	< 2e-16	-3.419	-1.495
σ^W_{ACC} * S5	26.264	1.482	< 2e-16	23.359	29.169
σ^W_{ACC} * S6	-4.417	0.484	< 2e-16	-5.366	-3.468
σ^W_{ACC} * S7	45.062	2.59	< 2e-16	39.986	50.138
σ^W_{ACC} * S8	-2.671	0.478	< 2e-16	-3.608	-1.734
σ^W_{ACC} * S9	1.091	0.622	0.079	-0.128	2.31
σ^W_{ACC} * S11	-2.302	0.483	< 2e-16	-3.249	-1.355
σ^W_{ACC} * S13	-2.119	0.475	< 2e-16	-3.05	-1.188
σ^W_{ACC} * S14	0.902	0.559	0.107	-0.194	1.998
σ^W_{ACC} * S15	-2.553	0.48	< 2e-16	-3.494	-1.612
σ^W_{ACC} * S16	-2.554	0.508	< 2e-16	-3.55	-1.558
σ^W_{ACC} * S17	-2.787	0.479	< 2e-16	-3.726	-1.848

Table 5: Coefficients obtained with Model 2 (Chest and Wrist Model, Data Sampled at 4Hz)

Coefficient	Estimate	Std. Error	p-value	Lower CI	Upper CI
(Intercept)	5.596	0.115	< 2e-16	5.371	5.821
σ^C_{EDA}	1.198	0.058	< 2e-16	1.084	1.312
σ^W_{EDA}	8.779	0.148	< 2e-16	8.489	9.069
σ^C_T	0.171	0.019	< 2e-16	0.134	0.208
min^W_T	-1.801	0.05	< 2e-16	-1.899	-1.703
min^C_{EMG}	-0.143	0.021	< 2e-16	-0.184	-0.102
σ_{inhale}	0.573	0.026	< 2e-16	0.522	0.624
σ_{exhale}	-0.266	0.021	< 2e-16	-0.307	-0.225
max_{breath}	0.636	0.017	< 2e-16	0.603	0.669
ie_{ratio}	-0.289	0.019	< 2e-16	-0.326	-0.252
μ^C_{ACC}	0.039	0.025	0.127	-0.01	0.088
σ^C_{ACC}	0.802	0.028	< 2e-16	0.747	0.857
σ^W_{ACC}	0.388	0.036	< 2e-16	0.317	0.459
σ^W_{HR}	0.209	0.024	< 2e-16	0.162	0.256
σ^W_{HRV}	0.173	0.02	< 2e-16	0.134	0.212
rms^W_{HRV}	0.282	0.057	< 2e-16	0.17	0.394
σ^C_{HR}	-0.74	0.095	< 2e-16	-0.926	-0.554
rms^C_{HRV}	0.061	0.02	0.002	0.022	0.1
σ^W_{EDA} * min^W_T	-4.665	0.08	< 2e-16	-4.822	-4.508
Participant					
S10	(Reference)				
S2	-0.171	0.116	0.14	-0.398	0.056
S3	-2.475	0.165	< 2e-16	-2.798	-2.152
S4	-1.63	0.213	< 2e-16	-2.047	-1.213
S5	-2.028	0.156	< 2e-16	-2.334	-1.722
S6	-4.503	0.152	< 2e-16	-4.801	-4.205
S7	-3.272	0.143	< 2e-16	-3.552	-2.992
S8	-1.01	0.12	< 2e-16	-1.245	-0.775
S9	0.308	0.115	0.007	0.083	0.533
S11	-3.188	0.142	< 2e-16	-3.466	-2.91
S13	-4.831	0.178	< 2e-16	-5.18	-4.482
S14	0.007	0.115	0.951	-0.218	0.232
S15	-0.115	0.149	0.443	-0.407	0.177
S16	-0.64	0.158	< 2e-16	-0.95	-0.33
S17	0.397	0.116	0.001	0.17	0.624
rms^W_{HRV} *** Participant**					
rms^W_{HRV} * S10	(Reference)				
rms^W_{HRV} * S2	-0.284	0.077	< 2e-16	-0.435	-0.133
rms^W_{HRV} * S3	0.542	0.095	< 2e-16	0.356	0.728
rms^W_{HRV} * S4	-0.424	0.254	0.095	-0.922	0.074
rms^W_{HRV} * S5	-0.561	0.16	< 2e-16	-0.875	-0.247
rms^W_{HRV} * S6	-0.3	0.094	0.001	-0.484	-0.116
rms^W_{HRV} * S7	-1.794	0.11	< 2e-16	-2.01	-1.578
rms^W_{HRV} * S8	-0.192	0.09	0.032	-0.368	-0.016
rms^W_{HRV} * S9	-0.358	0.078	< 2e-16	-0.511	-0.205
rms^W_{HRV} * S11	1.116	0.092	< 2e-16	0.936	1.296
rms^W_{HRV} * S13	-1.357	0.181	< 2e-16	-1.712	-1.002
rms^W_{HRV} * S14	-0.309	0.071	< 2e-16	-0.448	-0.17
rms^W_{HRV} * S15	-1.252	0.08	< 2e-16	-1.409	-1.095
rms^W_{HRV} * S16	0.089	0.092	0.331	-0.091	0.269
rms^W_{HRV} * S17	-1.226	0.075	< 2e-16	-1.373	-1.079

Table 6: Coefficients obtained with Model 2 (Chest and Wrist Model, Data Sampled at 4Hz)

Coefficient	Estimate	Std. Error	p-value	Lower CI	Upper CI
σ_{HR}^{C} * **Participant**					
σ_{HR}^{C} * S10	(Reference)			NA	NA
σ_{HR}^{C} * S2	0.213	0.117	0.07	-0.01632	0.44232
σ_{HR}^{C} * S3	0.276	0.111	0.013	0.05844	0.49356
σ_{HR}^{C} * S4	0.725	0.129	$<$ 2e-16	0.47216	0.97784
σ_{HR}^{C} * S5	-0.04	0.13	0.757	-0.2948	0.2148
σ_{HR}^{C} * S6	0.902	0.131	$<$ 2e-16	0.64524	1.15876
σ_{HR}^{C} * S7	0.561	0.126	$<$ 2e-16	0.31404	0.80796
σ_{HR}^{C} * S8	-0.629	0.112	$<$ 2e-16	-0.84852	-0.40948
σ_{HR}^{C} * S9	0.309	0.11	0.005	0.0934	0.5246
σ_{HR}^{C} * S11	1.714	0.125	$<$ 2e-16	1.469	1.959
σ_{HR}^{C} * S13	0.655	0.137	$<$ 2e-16	0.38648	0.92352
σ_{HR}^{C} * S14	0.377	0.105	$<$ 2e-16	0.1712	0.5828
σ_{HR}^{C} * S15	0.29	0.113	0.01	0.06852	0.51148
σ_{HR}^{C} * S16	0.548	0.11	$<$ 2e-16	0.3324	0.7636
σ_{HR}^{C} * S17	1.168	0.106	$<$ 2e-16	0.96024	1.37576

Table 7: Coefficients obtained with Model 3 (Wrist Model, Data Sampled using 60s Window Shift)

Coefficient	Estimate	Std. Error	p-value	Lower CI	Upper CI
(Intercept)	4.95	0.283	< 2e-16	4.395	5.505
σ^W_{EDA}	4.052	0.065	< 2e-16	3.925	4.179
σ^W_{ACC}	3.802	0.469	< 2e-16	2.883	4.721
σ^W_{HRV}	-0.001	0.017	0.937	-0.034	0.032
σ^W_{EDA} * σ^W_{ACC}	-2.03	0.046	< 2e-16	-2.12	-1.94
σ^W_{EDA} * σ^W_{HRV}	-0.113	0.027	< 2e-16	-0.166	-0.06
σ^W_{ACC} * σ^W_{HRV}	-0.095	0.029	0.001	-0.152	-0.038
Participant					
S10	(Reference)				
S2	-2.161	0.289	< 2e-16	-2.727	-1.595
S3	-2.572	0.29	< 2e-16	-3.14	-2.004
S4	-3.187	0.289	< 2e-16	-3.753	-2.621
S5	15.567	1.117	< 2e-16	13.378	17.756
S6	-5.842	0.295	< 2e-16	-6.42	-5.264
S7	47.569	3.114	< 2e-16	41.466	53.672
S8	-2.261	0.286	< 2e-16	-2.822	-1.7
S9	0.951	0.367	0.01	0.232	1.67
S11	-3.213	0.29	< 2e-16	-3.781	-2.645
S13	-9.277	0.308	< 2e-16	-9.881	-8.673
S14	-0.189	0.319	0.553	-0.814	0.436
S15	-2.282	0.287	< 2e-16	-2.845	-1.719
S16	-1.996	0.295	< 2e-16	-2.574	-1.418
S17	-1.865	0.286	< 2e-16	-2.426	-1.304
σ^W_{ACC} * **Participant**					
σ^W_{ACC} * S10	(Reference)				
σ^W_{ACC} * S2	-3.272	0.488	< 2e-16	-4.228	-2.316
σ^W_{ACC} * S3	-2.389	0.483	< 2e-16	-3.336	-1.442
σ^W_{ACC} * S4	-3.177	0.483	< 2e-16	-4.124	-2.23
σ^W_{ACC} * S5	29.137	1.962	< 2e-16	25.291	32.983
σ^W_{ACC} * S6	-4.884	0.491	< 2e-16	-5.846	-3.922
σ^W_{ACC} * S7	83.337	5.247	< 2e-16	73.053	93.621
σ^W_{ACC} * S8	-3.652	0.473	< 2e-16	-4.579	-2.725
σ^W_{ACC} * S9	0.953	0.612	0.12	-0.247	2.153
σ^W_{ACC} * S11	-2.257	0.479	< 2e-16	-3.196	-1.318
σ^W_{ACC} * S13	-0.038	0.486	0.938	-0.991	0.915
σ^W_{ACC} * S14	-0.267	0.547	0.625	-1.339	0.805
σ^W_{ACC} * S15	-3.42	0.475	< 2e-16	-4.351	-2.489
σ^W_{ACC} * S16	-3.241	0.504	< 2e-16	-4.229	-2.253
σ^W_{ACC} * S17	-3.652	0.475	< 2e-16	-4.583	-2.721

Table 8: Coefficients obtained with Model 4 (Chest and Wrist Model, Data Sampled using 60s Window Shift)

Coefficient	Estimate	Std. Error	p-value	Lower CI	Upper CI
(Intercept)	5.768	0.097	< 2e-16	5.578	5.958
σ^C_{EDA}	2.36	0.078	< 2e-16	2.207	2.513
σ^W_{EDA}	7.823	0.135	< 2e-16	7.558	8.088
σ^C_T	0.145	0.031	< 2e-16	0.084	0.206
min^W_T	-1.81	0.055	< 2e-16	-1.918	-1.702
min^C_{EMG}	0.042	0.027	0.126	-0.011	0.095
σ_{inhale}	1.769	0.048	< 2e-16	1.675	1.863
σ_{exhale}	-0.95	0.032	< 2e-16	-1.013	-0.887
max_{breath}	0.505	0.02	< 2e-16	0.466	0.544
ie_{ratio}	-0.214	0.023	< 2e-16	-0.259	-0.169
μ^C_{ACC}	0.472	0.029	< 2e-16	0.415	0.529
σ^C_{ACC}	0.918	0.036	< 2e-16	0.847	0.989
σ^W_{ACC}	0.302	0.044	< 2e-16	0.216	0.388
σ^W_{HR}	0.226	0.029	< 2e-16	0.169	0.283
σ^W_{HRV}	0.033	0.022	0.134	-0.01	0.076
rms^W_{HRV}	0.009	0.026	0.719	-0.042	0.06
σ^C_{HR}	-0.336	0.022	< 2e-16	-0.379	-0.293
rms^C_{HRV}	-0.041	0.018	0.023	-0.076	-0.006
$\sigma^W_{EDA} * min^W_T$	-3.919	0.07	< 2e-16	-4.056	-3.782
Participant					
S10	(Reference)				
S2	0.117	0.086	0.172	-0.052	0.286
S3	-1.362	0.15	< 2e-16	-1.656	-1.068
S4	-1.806	0.129	< 2e-16	-2.059	-1.553
S5	-3.272	0.138	< 2e-16	-3.542	-3.002
S6	-7.004	0.157	< 2e-16	-7.312	-6.696
S7	-3.28	0.121	< 2e-16	-3.517	-3.043
S8	0.452	0.104	< 2e-16	0.248	0.656
S9	1.228	0.097	< 2e-16	1.038	1.418
S11	-1.486	0.109	< 2e-16	-1.7	-1.272
S13	-6.052	0.173	< 2e-16	-6.391	-5.713
S14	1.037	0.102	< 2e-16	0.837	1.237
S15	0.213	0.145	0.143	-0.071	0.497
S16	0.533	0.157	0.001	0.225	0.841
S17	0.224	0.094	0.018	0.04	0.408

Table 9: Coefficients obtained with Model 1 - Quantifying Hetereogeneity

Term	Log Odds	Std. Error	Odds	Lower CI	Upper CI
Participant					
S2	-1.479	0.242	2.280000e-01	1.420000e-01	3.660000e-01
S3	-2.149	0.241	1.170000e-01	7.300000e-02	1.870000e-01
S4	-2.279	0.244	1.020000e-01	6.300000e-02	1.650000e-01
S5	12.174	0.736	1.936412e+05	4.580491e+04	8.186222e+05
S6	-4.238	0.245	1.400000e-02	9.000000e-03	2.300000e-02
S7	20.392	1.270	7.178041e+08	5.955920e+07	8.650935e+09
S8	-1.560	0.241	2.100000e-01	1.310000e-01	3.370000e-01
S9	0.656	0.312	1.926000e+00	1.046000e+00	3.548000e+00
S11	-2.505	0.243	8.200000e-02	5.100000e-02	1.320000e-01
S13	-5.570	0.250	4.000000e-03	2.000000e-03	6.000000e-03
S14	-0.034	0.270	9.660000e-01	5.690000e-01	1.640000e+00
S15	-1.548	0.242	2.130000e-01	1.320000e-01	3.420000e-01
S16	-1.537	0.249	2.150000e-01	1.320000e-01	3.500000e-01
S17	-1.362	0.240	2.560000e-01	1.600000e-01	4.100000e-01
σ^W_{ACC} *** Participant**					
σ^W_{ACC} * S2	-2.561	0.489	7.700000e-02	3.000000e-02	2.020000e-01
σ^W_{ACC} * S3	-2.686	0.477	6.800000e-02	2.700000e-02	1.740000e-01
σ^W_{ACC} * S4	-2.457	0.491	8.600000e-02	3.300000e-02	2.250000e-01
σ^W_{ACC} * S5	26.264	1.482	2.548917e+11	1.395468e+10	4.655768e+12
σ^W_{ACC} * S6	-4.417	0.484	1.200000e-02	5.000000e-03	3.100000e-02
σ^W_{ACC} * S7	45.062	2.590	3.716331e+19	2.318608e+17	5.956642e+21
σ^W_{ACC} * S8	-2.671	0.478	6.900000e-02	2.700000e-02	1.760000e-01
σ^W_{ACC} * S9	1.091	0.622	2.979000e+00	8.810000e-01	1.007400e+01
σ^W_{ACC} * S11	-2.302	0.483	1.000000e-01	3.900000e-02	2.580000e-01
σ^W_{ACC} * S13	-2.119	0.475	1.200000e-01	4.700000e-02	3.050000e-01
σ^W_{ACC} * S14	0.902	0.559	2.465000e+00	8.240000e-01	7.377000e+00
σ^W_{ACC} * S15	-2.553	0.480	7.800000e-02	3.000000e-02	1.990000e-01
σ^W_{ACC} * S16	-2.554	0.508	7.800000e-02	2.900000e-02	2.100000e-01
σ^W_{ACC} * S17	-2.787	0.479	6.200000e-02	2.400000e-02	1.580000e-01

Table 10: Coefficients obtained with Model 2 - Quantifying Hetereogeneity

Term	Log Odds	Std. Error	Odds	Lower CI	Upper CI
Participant					
S2	-0.171	0.116	0.843	0.671	1.058
S3	-2.475	0.165	0.084	0.061	0.116
S4	-1.630	0.213	0.196	0.129	0.297
S5	-2.028	0.156	0.132	0.097	0.179
S6	-4.503	0.152	0.011	0.008	0.015
S7	-3.272	0.143	0.038	0.029	0.050
S8	-1.010	0.120	0.364	0.288	0.461
S9	0.308	0.115	1.360	1.086	1.703
S11	-3.188	0.142	0.041	0.031	0.054
S13	-4.831	0.178	0.008	0.006	0.011
S14	0.007	0.115	1.007	0.803	1.263
S15	-0.115	0.149	0.892	0.665	1.195
S16	-0.640	0.158	0.527	0.387	0.718
S17	0.397	0.116	1.487	1.185	1.867
rms^W_{HRV} *** Participant**					
rms^W_{HRV} * S2	-0.284	0.077	0.753	0.647	0.876
rms^W_{HRV} * S3	0.542	0.095	1.719	1.427	2.072
rms^W_{HRV} * S4	-0.424	0.254	0.654	0.398	1.076
rms^W_{HRV} * S5	-0.561	0.160	0.571	0.417	0.781
rms^W_{HRV} * S6	-0.300	0.094	0.740	0.616	0.891
rms^W_{HRV} * S7	-1.794	0.110	0.166	0.134	0.206
rms^W_{HRV} * S8	-0.192	0.090	0.826	0.693	0.984
rms^W_{HRV} * S9	-0.358	0.078	0.699	0.600	0.815
rms^W_{HRV} * S11	1.116	0.092	3.054	2.551	3.655
rms^W_{HRV} * S13	-1.357	0.181	0.258	0.181	0.367
rms^W_{HRV} * S14	-0.309	0.071	0.734	0.639	0.843
rms^W_{HRV} * S15	-1.252	0.080	0.286	0.244	0.334
rms^W_{HRV} * S16	0.089	0.092	1.094	0.913	1.310
rms^W_{HRV} * S17	-1.226	0.075	0.293	0.253	0.340
σ^C_{HR} *** Participant**					
σ^C_{HR} * S2	0.213	0.117	1.238	0.983	1.558
σ^C_{HR} * S3	0.276	0.111	1.318	1.060	1.638
σ^C_{HR} * S4	0.725	0.129	2.065	1.603	2.659
σ^C_{HR} * S5	-0.040	0.130	0.961	0.745	1.239
σ^C_{HR} * S6	0.902	0.131	2.465	1.908	3.183
σ^C_{HR} * S7	0.561	0.126	1.752	1.369	2.244
σ^C_{HR} * S8	-0.629	0.112	0.533	0.428	0.664
σ^C_{HR} * S9	0.309	0.110	1.363	1.098	1.690
σ^C_{HR} * S11	1.714	0.125	5.552	4.349	7.088
σ^C_{HR} * S13	0.655	0.137	1.924	1.471	2.517
σ^C_{HR} * S14	0.377	0.105	1.458	1.187	1.792
σ^C_{HR} * S15	0.290	0.113	1.337	1.070	1.669
σ^C_{HR} * S16	0.548	0.110	1.729	1.395	2.144
σ^C_{HR} * S17	1.168	0.106	3.215	2.610	3.959

Table 11: VIF Values - Model 1

Term	VIF
σ^W_{EDA}	3.399
σ^W_{ACC}	256.719
σ^W_{HRV}	15.832
$\sigma^W_{EDA} * \sigma^W_{ACC}$	2.657
$\sigma^W_{EDA} * \sigma^W_{HRV}$	10.747
$\sigma^W_{ACC} * \sigma^W_{HRV}$‘	9.249
S2	34.238
S3	26.465
S4	24.095
S5	105.757
S6	21.026
S7	247.069
S8	30.031
S9	56.252
S11	21.161
S13	14.416
S14	34.539
S15	32.267
S16	35.668
S17	34.546
$\sigma^W_{ACC} *$ S2	26.664
$\sigma^W_{ACC} *$ S3	42.962
$\sigma^W_{ACC} *$ S4	20.439
$\sigma^W_{ACC} *$ S5	105.100
$\sigma^W_{ACC} *$ S6	23.205
$\sigma^W_{ACC} *$ S7	246.714
$\sigma^W_{ACC} *$ S8	38.854
$\sigma^W_{ACC} *$ S9	56.521
$\sigma^W_{ACC} *$ S11	25.507
$\sigma^W_{ACC} *$ S13	73.648
$\sigma^W_{ACC} *$ S14	28.493
$\sigma^W_{ACC} *$ S15	41.220
$\sigma^W_{ACC} *$ S16	26.571
$\sigma^W_{ACC} *$ S17	36.072

Table 12: VIF Values - Model 2

Term	VIF
σ^{C}_{EDA}	1.727
σ^{W}_{EDA}	8.985
σ^{C}_{T}	1.699
min^{W}_{T}	13.052
min^{C}_{EMG}	1.336
σ_{inhale}	1.562
σ_{exhale}	1.680
max_{breath}	1.238
ie_{ratio}	1.530
μ^{C}_{ACC}	2.551
σ^{C}_{ACC}	1.713
σ^{W}_{ACC}	1.546
σ^{W}_{HR}	2.780
σ^{W}_{HRV}	1.791
rms^{W}_{HRV}	15.271
σ^{C}_{HR}	37.769
rms^{C}_{HRV}	1.859
σ^{W}_{EDA} * min^{W}_{T}	11.358
S2	6.115
S3	6.461
S4	7.441
S5	6.633
S6	3.540
S7	2.442
S8	4.281
S9	5.542
S11	3.831
S13	5.346
S14	5.590
S15	6.011
S16	8.148
S17	5.366
rms^{W}_{HRV} * S2	2.299
rms^{W}_{HRV} * S3	3.608
rms^{W}_{HRV} * S4	5.803
rms^{W}_{HRV} * S5	4.010
rms^{W}_{HRV} * S6	1.908
rms^{W}_{HRV} * S7	1.555
rms^{W}_{HRV} * S8	1.974
rms^{W}_{HRV} * S9	2.253
rms^{W}_{HRV} * S11	2.447
rms^{W}_{HRV} * S13	3.145
rms^{W}_{HRV} * S14	2.970
rms^{W}_{HRV} * S15	2.156
rms^{W}_{HRV} * S16	3.948
rms^{W}_{HRV} * S17	3.040

Table 13: VIF Values - Model 2

Term	VIF
σ_{HR}^{C} * S2	3.718
σ_{HR}^{C} * S3	4.708
σ_{HR}^{C} * S4	2.297
σ_{HR}^{C} * S5	2.953
σ_{HR}^{C} * S6	2.460
σ_{HR}^{C} * S7	2.827
σ_{HR}^{C} * S8	3.962
σ_{HR}^{C} * S9	4.557
σ_{HR}^{C} * S11	2.883
σ_{HR}^{C} * S13	2.512
σ_{HR}^{C} * S14	5.135
σ_{HR}^{C} * S15	3.782
σ_{HR}^{C} * S16	3.906
σ_{HR}^{C} * S17	4.722