Portfolio 7: Signe Kløve and Thea

Link to code:

<https://github.com/YpipY/EXPMETH3/blob/master/Assignment%204/newdata_sandt.Rmd>

1) How do you preprocess heart rate and respiration data?

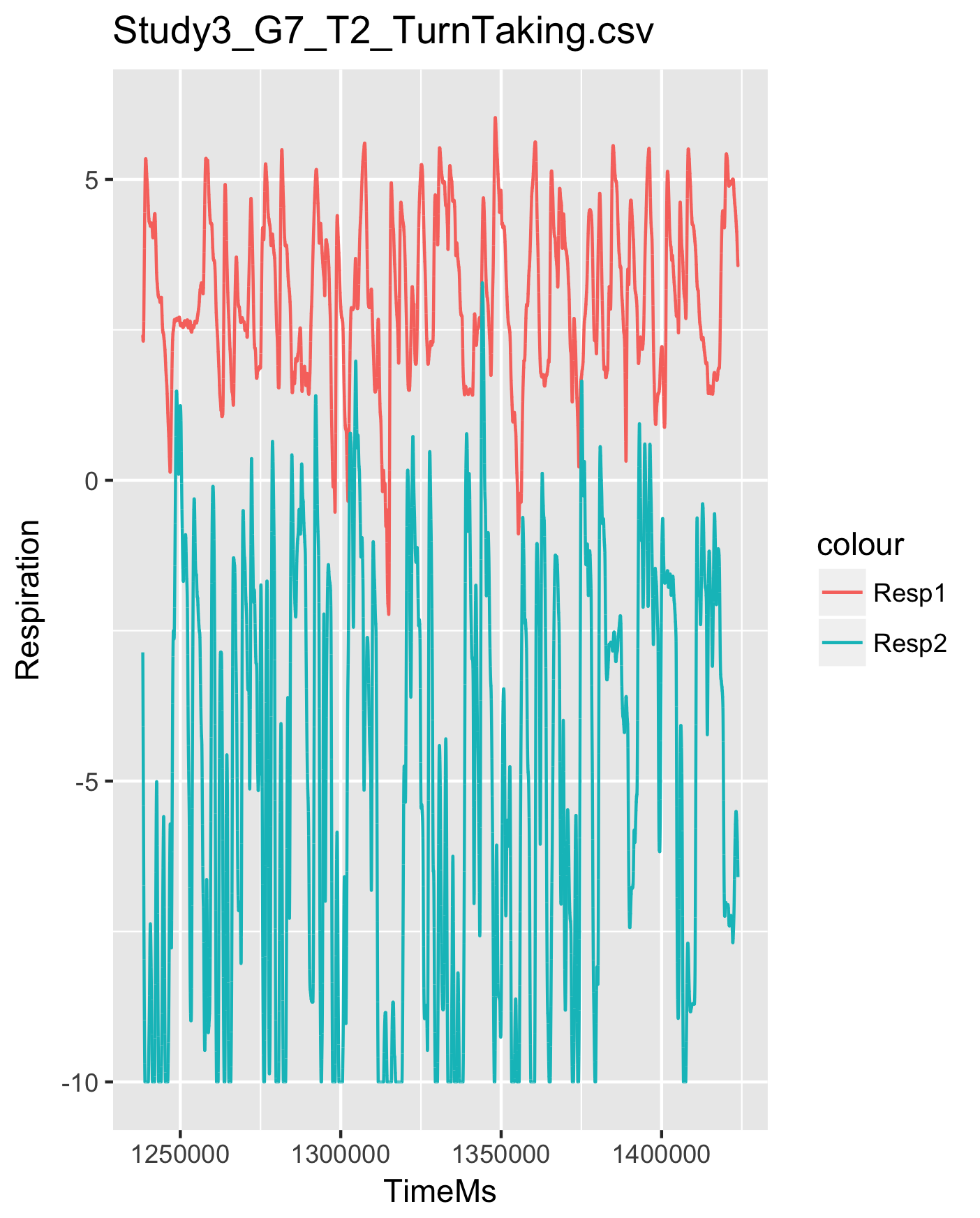
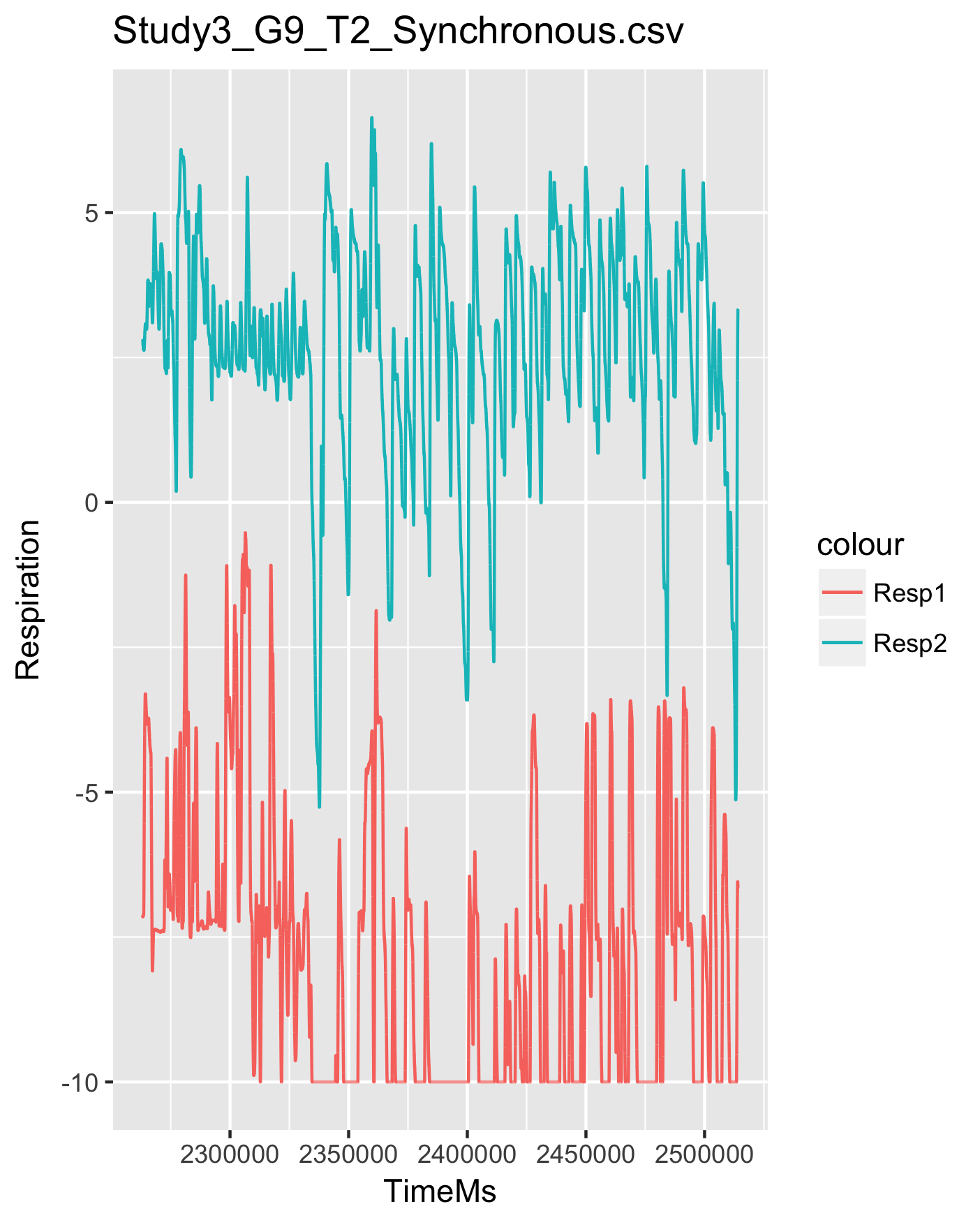
**Describe the process. If any data needs to be excluded, list the excluded data and motivate the exclusion.**

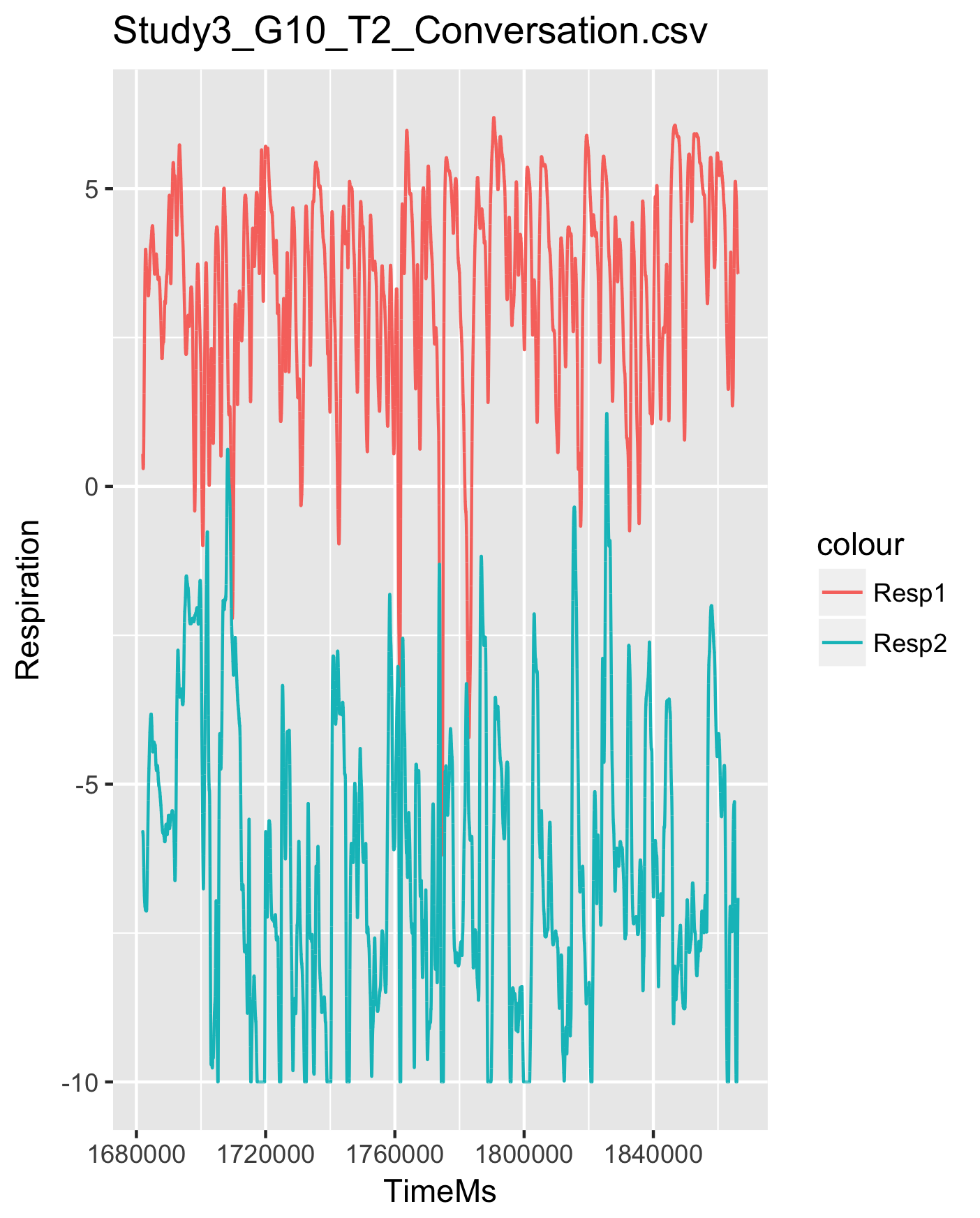
**The process:**

**Plot of raw data and exclusion**

We plotted the data; one plot for respiration (RE) and one plot for heart rate (HR) for each pair of participants. We looked at the plots, to see if any should be excluded from the analysis. By visually inspecting the HR-plots we decided to keep them all.  
The following trials had some artifacts in the RE data: *G7\_T2\_TurnTaking, G9\_T2\_Synchronous, G10\_T3\_TurnTaking, G10\_T2\_Conversation*. The reason for these artifacts is probably due to the belt not being tightly fixed around the participant, when not measuring anything it probably just records -10. The optimal solution would be to exclude the data for these trials, but for this first analysis we did not exclude them.

**Plot of the raw data for the 4 trials, which we considered to exclude:**





**Downsampling of data**

Due to computation time we need to downsample the data. We found the mean for each 100 datapoints, then again for the next 100 datapoints and so on...

A better approach would be to use a “sliding mean” approach, where the mean is found for the first 200 datapoints and then “slide” 100 datapoints and then find the mean for these 200 datapoints, then slide 100 datapoints, and so on… In this way we would keep some of the information from the previous section. We know it is a continuous process developing over time and therefore we expect shared information between intervals.

**Outlier removal**

In order to remove outliers we decided to keep all data within a range of the mean plus or minus a specified threshold (2.5) times the standard deviation (2.5\*sd).

We replaced the values outside this range with the mean value for the time series.

If there is a lot of outliers in both time series and we replace all of them with the mean, then it might distort the rcqa analysis, as it will look like recurrence between the two time series, then in fact it’s just due to a substitution with the mean in the two time series.

In order to control for this, it would have been a good idea to look at how many outliers was removed for each pair and then exclude them if a certain threshold were reached.

Thus, out approach can have caused problems and disturbed the coordination analysis. A better approach would be to use a cubic spline (and we will do that next time...)

**Scaling**

We scaled the data by z-scoring the time series for HR and for RE. This was done in order to be able to compare between the participants, as one person might have a lower or higher baseline compared to the other.

**Plot of preprocessed data**

We plotted the preprocessed data to eyeball whether the preprocessing were sufficient. It was okay.

2) Do you observe interpersonal coordination in heart rate and respiration?

**Describe your control baseline, the method used to quantify coordination, and the statistical models used to infer whether coordination was higher than in the baseline. Report the results of the models.**

**Description of control baselines:**

Control baselines:

In order to assess whether coordination of HR and RE is due to people simply performing the same task or if it actually is due to the fact that they are performing the task in the presence of each other, we need a baseline to compare with. We used two different kinds:

*Shuffled pairs:* in the shuffled control baseline, all the recordings are shuffled, i.e. the order of each participants timeseries is completely randomized. We used the sample() function to execute this.

*Surrogate pairs:* in surrogate pairs, all possible 153 combinations of pairs were joined. In case that the real pairs were not coordinated significantly more than the surrogate pairs, it would be an indicator of the task causing the coordination and not the social interaction in that condition (synchronous, selfpaced or turntaking).

**Methods to quantify coordination:**

In order to quantify coordination we used different output parameters from the RCQA-analysis:

* **Recurrence rate (RR):** how many recurrences in percent.
* **Trapping time (TT):** vertical lines
* **Mean length of lines (L):** at least two dots to constitute a line
* **Length of the longest line (maxL)**
* **Determinism (DET):** the ratio of diagonal lines of all the recurrences, i.e. how many of the recurrences are a part of a diagonal line

Using the OptimizaParam-function we found optimal parameters for HR and RE in the rcqa analysis:

* HR: delay: 4 embed : 20 radius: 2.33
* RE: delay: 31 embed : 2 radius: 0.38

**Statistical models**

The following model was used to infer whether coordination was higher than in the baseline: ***model = lm(parameter ~ Condition + baseline, data)***

**Results of the models:**

For all the parameters *baseline shuffle* is significantly lower than *real pairs* in both heart rate and respiration. For *baseline surrogate* compared to *real pairs* we found no significant difference across all parameters in both HR and RE.

When creating this model, we are looking across conditions. This might be problematic since we would assume different differences. E.g. in the *surrogate baseline* we would expect to see one difference in the parameters between the conditions synchronous and conversation and another difference between synchronous and conversation for *real pairs*. You could argue that the synchronous condition is more task driven and that it does not matter whether the participants are actually doing the task in the presence of each other. Also, it could be argued that the conversation condition is more driven by, or dependent on, the participants being in the presense of each other, thus interacting socially. This way of analysis would result in a difference in how surrogate and real pairs affect the conditions and thereby possibly result in an interaction effect.

Thus, this type of relation would not become clear by looking at a model which only consider baselines, therefore it’s important to also look at the interaction between condition and baseline.

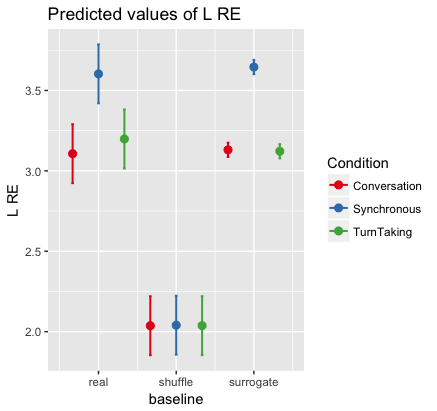
***interaction\_model = lm(parameter ~ Condition \* baseline , data)***

The only significant interaction effect was found in the following model:

***int\_L\_RE\_model = lm(L for RE ~ Condition \* baseline , data)***

The significant interaction effect was between Condition Synchronous and baseline shuffle (β= -0.49 (SE= 0.19), t = -2.64, p<.05), going from Condition Conversation and baseline real pairs.

Beneath is a plot of this interaction effect.



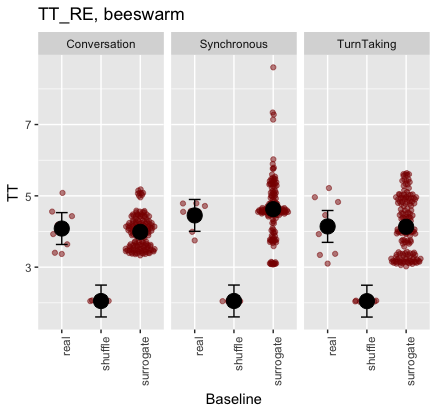
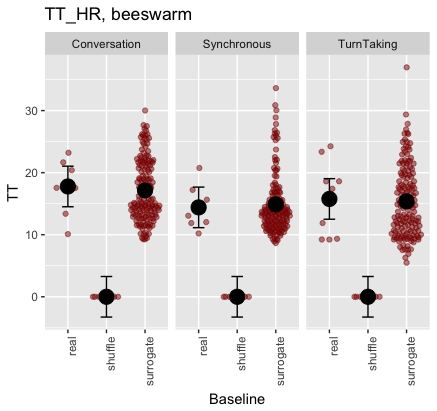
There is a significant interaction when we go from conversation to synchronous and from real to shuffle. It means that the two type of pairs have a different effect on the conversation and synchronous conditions. If we look at the plot, we can see that shuffle pairs has a somewhat equal mean length of recurrence in both the conversation and synchronous condition. This makes sense as the timeseries are completely shuffled. For the real pairs there is a difference between mean length of recurrense for synchonous and conversation. There is a higher mean length of recurrence for condition synchronous compared to conversation. This makes sense, as the participants in condition synchronous are “forced” to follow the same pace of speech production and therefore align their breath. We can see that the errorbars do not overlap in the real pairs and this makes it a significant interaction compared to some of the other plots, where we see a similar patterns of difference between conditions for real pairs and no difference between shuffle pairs.

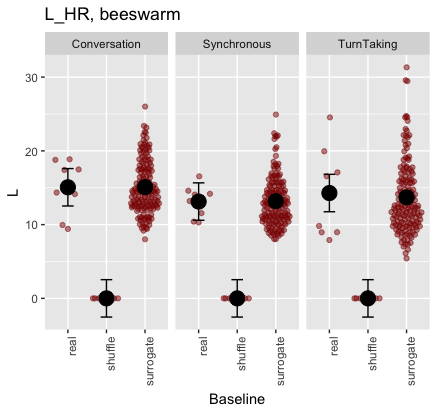
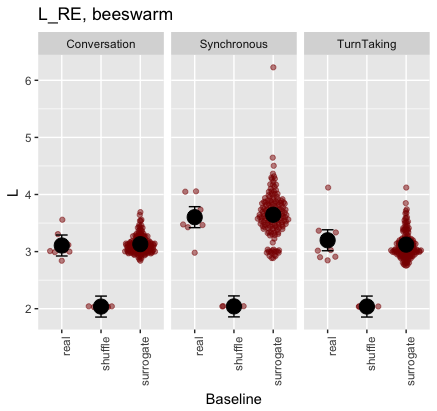
It can be discussed how much this interaction between real and shuffle inform us on the real cause of the differences between conditions. It is therefore important that we have included surrogate controls. If there is no difference between surrogate and real pairs, this would suggest that the coordination is purely due to performance of task, and not the social interaction.

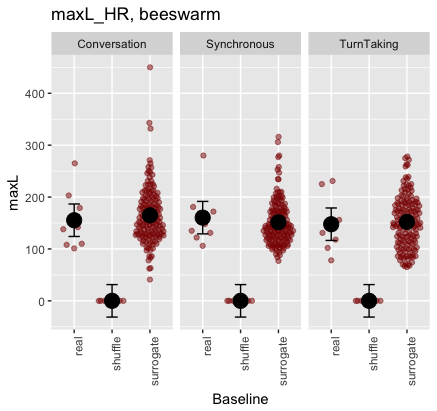
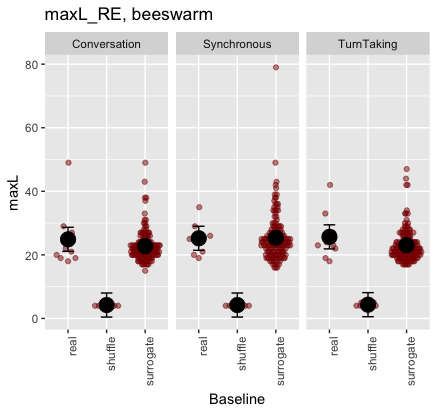
BEESWARM PLOTS OF MODEL’S PREDICTIONS

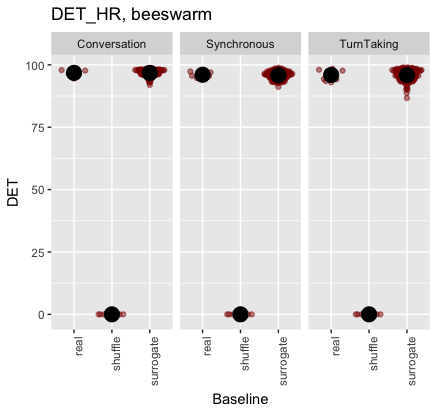
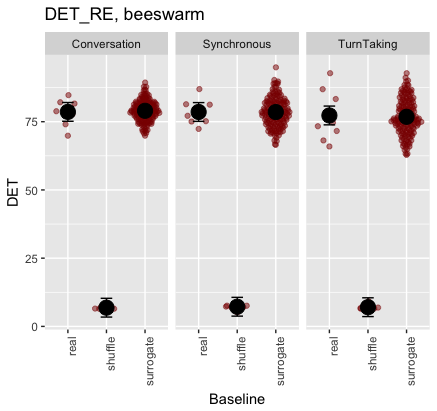
On the plots we can see how the parameters differ across the three conditions. In almost all the plots we see that the confidence intervals overlap across conditions when we look at the real pairs. This is also evident by the fact that we found no main effect of condition, when testing on the real pairs only. We also see an overlap in the confidence intervals between the real and the surrogate pairs. This indicates, as we also concluded from looking at the interaction model, that the coordination in recurrence might be more driven by the task, than by the participants being in the presence of each other.

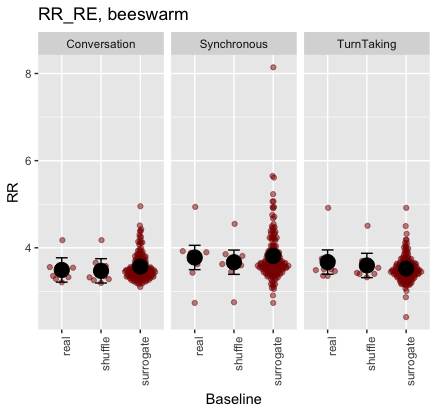
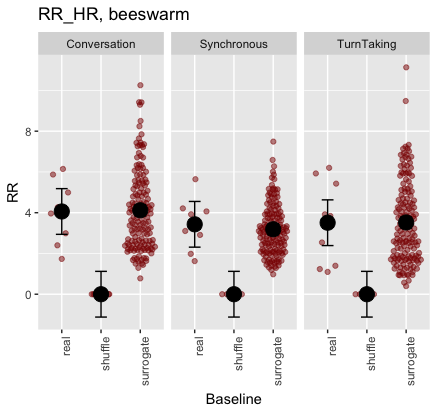
It is also clear from the beeswarm plots that we have more data points for the surrogate pairs (153). Therefore we see a much more narrow confidence interval surrounding the prediction. This interval is much wider for shuffle and real pairs, as we only have 10 data points. This also makes it difficult to find an effect of condition in real pairs, as we will easily get overlapping confidence intervals.











3) Do you observe differences in coordination between conditions?

**Report the models and results.**

This model was used to examine whether there is a difference in coordination between conditions.

***condition\_model = lm(parameter ~ Condition , data of real pairs)***

None of the recurrence measures is significantly predicted by Condition. Thus, there is no significant difference in the recurrence of HR or RE across conditions. Here we looked only at the parameters for the real pairs.

If we had found a significant effect of condition (e.g. higher recurrence of HR in conversation compared to turntaking), it would be important to check whether this effect is due to the pair of participants doing the task together, or if it’s mainly because of the kind of task. This is the reason we created a model with an interaction effect between condition and baseline.

4) Is respiration coordination a likely driver of heart rate coordination?

**Describe how you would test for it.**

One idea would be to create a model where parameters of respiration (e.g. RR) predicts parameters of heartrate. Thereby looking into whether coordination of respiration can predict coordination of heartrate. If it can, then it is could be argued that respiration is a driver for heart rate.