# Presentation of research questions

As previously introduced, this paper will investigate how well the model proposed by Maldonado et al. (2019) generalises to another area of interest in mouse-tracking research, more specifically the process of spoken-word recognition in relation to phonological similarity done by Spivey et al. (2005). The investigation will be done by replicating the experiment by Spivey et al. (2005) and applying the LDA model to the resulting data. We “hypothesise” that for the model to generalise well, it must perform accurate label classifications well above chance and perform close to the topline performance. Chance performance is defined as a model classifying random labels. The topline performance is defined as in Maldonado et al. (2019), where the topline is a model trained and tested on all data, creating a model overfitting the data and a resulting performance above what is expected. Performance of the models will be evaluated based on their AUC-ROC score.

Additionally, this paper will also investigate how well the model performs compared to traditional mouse-tracking measures ability to classify the data. The measures compared to the model will be area under the trajectory curve, maximal absolute deviation, x-flips, x-reversals, and average deviation.

We will also explore possible improvements of the model by an investigation of the optimal number of PCA-features compared to the 13 PCA features chosen by the original paper.

The results and possibilities of the model will be evaluated and discussed after reporting of the analysis.