

For this homework you will be separated into **5 groups**, each including **6 people**. You can form groups as you wish and you can also use the “Search for Teammates” feature on Piazza. **Each group will submit one report on eCampus.**

Objective

The goal of this homework is to design a data collection experiment that will explore human engagement. Engagement refers to the process of “starting and maintaining the perceived connection” between interacting individuals. It constitutes the basis for social connection and understanding and has been related—among others—to attentional involvement, emotional connectedness, and social rapport. Modeling engagement among others can be useful in health, education, entertainment applications, e.g. quantifying engagement levels of children with Autism during therapy, class engagement, or even people’s engagement when interacting with a robot or a conversational agent (e.g. Alexa). You will design a data collection experiment that aims to explore multimodal indices of engagement and to analyze the resulting data using statistical techniques.

Part 1: Designing the data collection

1. **(2.5 points)** Perform a short literature review on how engagement is perceived, elicited, and measured during dyadic interactions. You can find research papers from the ones that we have already discussed in class, as well as online. Please refer to the notes from **Margaret J. Foster on “Systematic Literature Review” (October 18th)**, as well as to a variety of online resources, e.g., <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3715443/>. Please describe your findings in 1-2 paragraphs in your report.
2. **(2.5 points)** Based on what you have read, plan a data collection experiment to study engagement in dyadic interactions or any other types of task that would elicit different degrees of engagement.
 - (a) Think about how you can manipulate the experiments to elicit two different conditions: (i) no or low engagement and (ii) high engagement.
 - (b) Participants in the study will be your classmates or your friends, but **make sure that participants are not from your own group** (why?). You can ask participants to take surveys regarding anything you think that might be useful for your experiment.
 - (c) Think about the types of bio-behavioral markers that might reflect change in engagement based on your setup. You can have recordings related to audio (from your laptop or mobile device) and physiology (from your smartwatch or by borrowing the E4 devices from the lab). You can use one, some of them or all of them. **Hint:** You can use Audacity to record audio on your computer: <https://www.audacityteam.org/>
 - (d) Form a hypothesis regarding your experiment. Think about how the various bio-behavioral markers differ between the two experimental conditions.

- (e) Think the extent to which your design is representative and generalizable in the real-world (external validity) and the degree to which your results can be potentially attributed to the independent variable, i.e. the two different conditions of engagement (internal validity). Please refer to **Dr. Winfred's notes on research validity (September 6th)**.
- (f) **Be creative :)** If there is an additional resource that you might need, please let me know. If you would like to use the E4 devices, please let me know which dates you would like to use them, so that we can set up a schedule.

In the report, please describe the proposed data collection design explaining (i) the two different conditions and why can these result in potentially different engagement levels; (ii) the types of signals that you are going to examine and your rationale about choosing those; (iii) your hypothesis, i.e., what are the expected findings and why (e.g., “Our hypothesis is that during the high-engagement conversation individuals will exhibit higher values of fundamental frequency F0, because previous work [REF] has shown that high engagement has been related to high emotional arousal, which is captured through F0 measures.” or “Our hypothesis is that during high engagement individuals will exhibit more synchronized physiological patterns, since engaging interactions are more likely to elicit common physiological reactivity between people.”).

Part 2: Performing the data collection

(2.5 points) Based on the experimental design from part one, each team will collect data from **2 sessions: 1 session with high engagement and 1 session with low engagement**. Each session will have a duration of **1-3 mins**. Depending on your design, different individuals can participate in each session. If you would like to collect more sessions or have longer sessions, you are free to do so.

During the data collection:

1. You will give clear instructions to the participants on what they should do. **Make sure not to bias your participants with information that they do not need to know, i.e., don't disclose the hypothesis of your study.**
2. Make sure that you keep track of the start and end time of each session (and possibly the different parts of the session) and synchronize the different sources of data.

In the report, please describe how data collection was performed (e.g., type of equipment, environment conditions), the number of participants and their demographics, the duration of each session, and any other procedures that you might have used (e.g., how was synchronization achieved?).

Part 3: Analyzing the data

(2.5 points) Extract appropriate features from each signal, e.g., prosodic features from audio, average signal levels from EDA. Compare the values of the corresponding features between the two different engagement levels. In the report, please describe and discuss your results and whether they support your hypothesis. If not, this is totally fine, but please provide potential reasons why. **Hint:** If you would like to perform a statistical analysis, you will have to segment each session in smaller segments, otherwise, you will not have enough samples. You can decide either to do a statistical analysis between the multiple segments of each session, or just report the values of the features for the entire session.

(2.5 points) Individual bonus point: Human subjects protection training

An important component of working with human data is to be compliant with the federal and state regulations. CITI training offers general knowledge on these. As a bonus point, complete the CITI training found in the following link and send me your report: <http://rcb.tamu.edu/humansubjects/signing-up-for-human-subject-protection-training-in-citi-1>. The good news is that once you do a CITI training it is valid for a long time and can be used in any other project that you might need it.