**Title**: Social Interaction Detection

**Team Members**: Grace Lee (gwl442), Hsiao-Yuan Chen

**Description**:

**Training Data**: 15 to 20 per class, 3 to 5 minutes each

* A person talking alone (reading a passage)
* 2 or more people chatting
* Talking on the phone
* Talking to a child (we are unable to collect this ourselves)
* Giving a lecture (use Zoom videos from class)
* Eating together
* Arguing

As our team members are bilingual/trilingual, we aim to collect and create training data that encompass more than a single language.

**Motivation**:

The difficulty in the project lies in the fact that every one of these audios has people talking in it. Unlike Assignment 2, where the difference between each class was large, here the difference is way more subtle.

This is interesting because humans can easily recognize whether the audio is of 2 people chatting or of people arguing, but to our best knowledge, Voice Activity Detection (or Speech Detection) can only detect whether human voices are present.

**Contribution**:

* We propose to utilize audio (and audio only) to classify social interaction. Since only audio is needed (no visual recognition is involved), privacy is protected.
* We demonstrate the difference between a training set with only one language, and a richer training set with multiple languages.

**Description of previous research in the area** (with references)

In previous work, this problem was solved using audio alongside with other sensors. Their weakness, their strength. We are different because we used [device] blahblahblah

One shortcoming of next2me is that it is not designed for activities in short time( less than 5 minutes), but short time social interaction can be important sometimes, but using next2me, this kind of information may not be captured.

The advantage of it is that it works well even when the phone is in pocket. In addition, it use wifi and audio to fulfill its goal and these two resources are more privacy-sensitive than images. However, wifi may still leak some extra information. It may be better if we can make this through audio only.

**Resources you may need** (in terms of hardware, software, etc)

* Smart Phone

**Expected timeline**

*Today~11/1* Finish collecting all data (training + testing). We need at least 140 different audio recordings, so we need to collect an average of 14 per day.

*11/2~11/7* Finish first draft of the code

11/8~11/14 Experiment with different features. May need to collect more data if necessary.

11/15~11/24 Buffer

11/25~11/29 Work on presentation

11/30 Presentation

12/7 Web Page

**References**

* <https://dl.acm.org/doi/abs/10.1145/3144457.3144500?casa_token=RB-8RzBT0g4AAAAA:jX2n9SapWzGrMRBQDAs80MJSZVPvuTtoXhMK8LN2M4QohQmvDSs04kCMGfbO3PAB4WoJScPnE7WV>
* <https://ieeexplore.ieee.org/abstract/document/6639176?casa_token=BeH94nnlHh8AAAAA:TfXDY_f3O9iqFINdJKwVZpwlGlZKnVjHxalcIINBvd9ra9CYcFunvOrT6mFJQww_Vyd1b5B7MQ>
* <https://formative.jmir.org/2020/8/e18751/?utm_source=TrendMD&utm_medium=cpc&utm_campaign=JMIR_Formative_Research_TrendMD_0>
* <https://arxiv.org/pdf/1903.11269.pdf?fbclid=IwAR156ziVTi_w_SpeRpygd245vyRnFnGTd2L9T_8AlKdNOfHqWkNf8Q67Neg>
* <https://ieeexplore.ieee.org/document/9127390>
* <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2953803/>
* <http://www.cvc.uab.es/CVCRD2016/pdfcloud/files/aghaei.pdf>