A MACHINE LEARNING MODEL FOR LAUGHTER DETECTION

Meeting 28.04.28

Confusion matrices

- *"Each row shows the percentage of laughter segments classified as a certain subclass for a given threshold"* (from the thesis)
- **Correct:** Each row shows the percentage of all **predicted** laughter segments located in a certain subclass for a given threshold.
  - i.e., we look at all segments predicted as laughter and check their true subclass

Ethical concerns

- Shouldn't discriminate against a certain race
  - Cultural differences in laughter *(Truong, 2007)*

Features

- PLP (Perceptual Linear Prediction)
  - Extracted from spectrogram (like MFCCs)
  - To capture the spectral properties
- Modulation spectrum features
  - *"capture the rhythm and repetitive syllable sounds"* (Truong and van Leeuwen, 2007)
- RMS (Root Mean Squared Energy)
- AC PEAK
  - *"highest normalized cross correlation value found to determine F0"* (Knox, 2006)
- F0 (Fundamental frequency)
- Spatial cues
  - Cross-correlating the signal from several microphones
  - Used to capture the direction from which a signal comes

Prosodic Features

Features relating to rhythm and intonation, e.g.

- Pause
- Intonation
- Volume
- Tempo

IDEAS FROM MEETING

Possible other corpora

- https://www.colorado.edu/cwa/
  - From a conference where an unsupervised clustering algorithm already found a cluster for laughter
    - This suggests that laughter was clearly separable from other events

Linguistic nature of laughter

- What does laughter convey?
- Is ICSI a good match for the type of laughter we are looking for?

VAD and Wake Word Detection research

- Are we able to make use of the insights from these fields?
  - *Voice Activity Detection (VAD) has some similarities with our task*
    - e.g. not getting triggered by noise
  - *Wake word detection is a much larger research area (Siri, Alexa, etc.)*
    - Also deals with the end user's privacy concerns