Week 9 (October 28-November4)

Tools used:

Python

Goal:

Objective:

For Kaya’s grant, we need statistics to claim what we have analyzed, and what portion did we find accurate vs not accurate.

How much crying data do we have?

(%, duration, count)

1. Episodes created by Episode.py (Count, Duration,%)
2. Episodes above Max Duration Threshold (xVals)

How much audio data did I cover (categorized)

(duration)

4 second gap / 5 seconds minimal duration for each episode

Histogram of count of episodes & total duration of crying (hours) for raw (episode.py) and crying episode that passes threshold

Weighted average of accuracy after max threshold

Results:

DensityAnalyze

1. Dataframe structure and what each component means in processfile()
   1. Participant -> participant identification
   2. Total Duration –> Actual Duration Analyzed -> The duration of an episode (Will be displayed as the sum of all durations of episodes)
   3. Total Occurring Episodes -> The total number of episodes for each participant
   4. Total Categorized Episodes -> The number of episodes that I went through
   5. Total Categorized Episodes Duration Sum – Actual Duration Analyzed -> The duration of an episode that I went through (Will be displayed as the sum)
   6. Episode (Start to End) Duration -> The duration of an episode from the very first occurrence to the very last occurrence (including times in between that may not necessarily be an episode)
   7. Categorized Episode (Start to End) Duration -> the duration of an episode from the very occurrence to the very last occurrence (including times in between that may not necessarily be an episode) for episode that I categorized
   8. Episode Count above Max Threshold -> The number of episodes that were counted (above the indicated max duration threshold)
   9. Episode Duration above Max Threshold -> Actual Duration Analyzed -> The total duration of all episodes that I analyzed that passes the maximum duration of a vocalization
2. Method execute1(max) and get()
   1. Calculates weighted average of accuracy vs max vocalization duration. Performs calculation based on adding all occurrences of categorization above the max threshold and multiples the corresponding occurrence / total \* density at a certain x val. The total accuracy is the summation of all these.
3. Plot(df)
   1. You can choose which ones to make a histogram