Computational Human AI Lab (CHAI)

Emotion Regional Saliency in SER

Research project by Huan Zhang
Mentored by Aneesha Sampath
June ‘25 - Dec ‘25

TABLE OF CONTENTS

01

Introduction

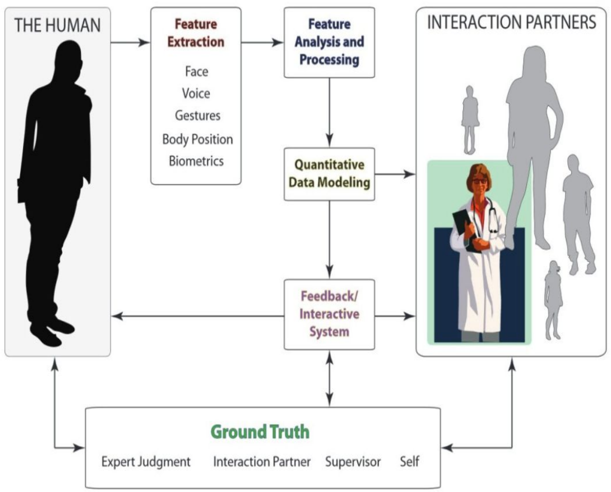
CHAI Lab

The CHAI Lab is directed by Prof. Emily Mower Provost, a professor of computer science at the University of Michigan.

The main goals of CHAI lab are to advance speech-centered machine learning for human behavior detection by focusing on three main areas:

1) Emotion Recognition
2) Mental Health Modeling
3) Assistive Technology for bipolar disorder and aphasia

In this lab, researchers are able to develop novel algorithms and machine learning tools while learning more about the underlying workings of human behavior.



Behavioral Signal Processing Approach to Modeling Human-centered Data

https://emp.engin.umich.edu/research

Project Goals & Vision

The overarching vision of this research project is to improve SER (speech emotion recognition) systems by making them more aware of where emotion occurs.

This is beneficial for downstream tasks like mental health monitoring, creating more assistive chatbots, and more by localizing exact points in speech where emotion is salient.

Throughout this project, the main milestones include 1) Discovering interesting patterns in the data/annotations of the MSP-Podcast, and 2) Discovering how data patterns relate to saliency maps. The main goal is to discover a novel method to address saliency gaps.

Emily Mower Provost
Professor, Computer Science and Engineering

Aneesha Sampath
Graduate Student

Huan Zhang
Undergraduate Student

02

Data Exploration

Overview & Tasks

Main Goal

Learn the landscape of the MSP-Podcast & establish a performance benchmark

Data

Create CSV with data separated into 9 bins for activation and valence

Binning

Listen to samples in each bin and note any patterns in acoustics or language

EDA

Create table & bar graph showing number of samples per bin

Plots

Create radar plots showcasing the emotion associated with each of the 9 bins

Findings

Listen to samples for each classification of emotion and note any unexpected patterns

Risks

Analyze outcomes and predict challenges the model may encounter

Skills Learned

- Creating CSVs using Pandas Dataframes
- Creating radar plots using Plotly
- Navigating MSP-Podcast audios and data files
- Setting up environment in VSCode and cloning Github repositories

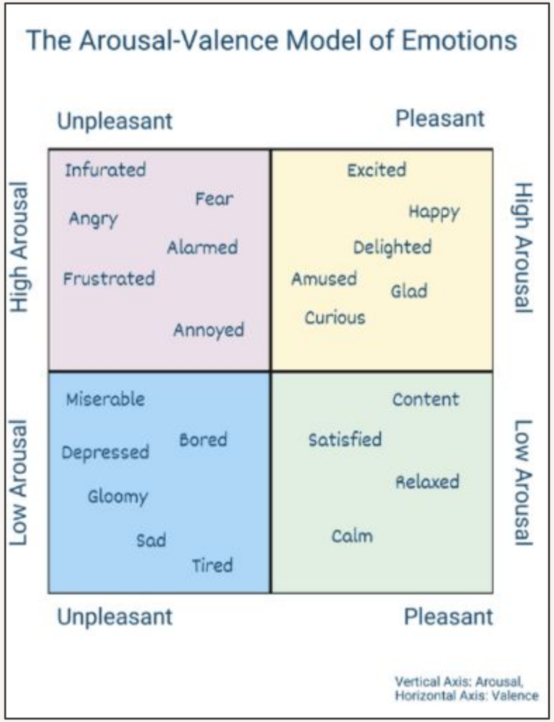
Activation & Valence

Activation (or arousal):

- Definition: The degree of physiological and psychological energy associated with an emotion.
- Range: Low activation (calmness, relaxedness), to high activation (excitement, energy)
- Examples: Calm or lethargic versus excited or alert

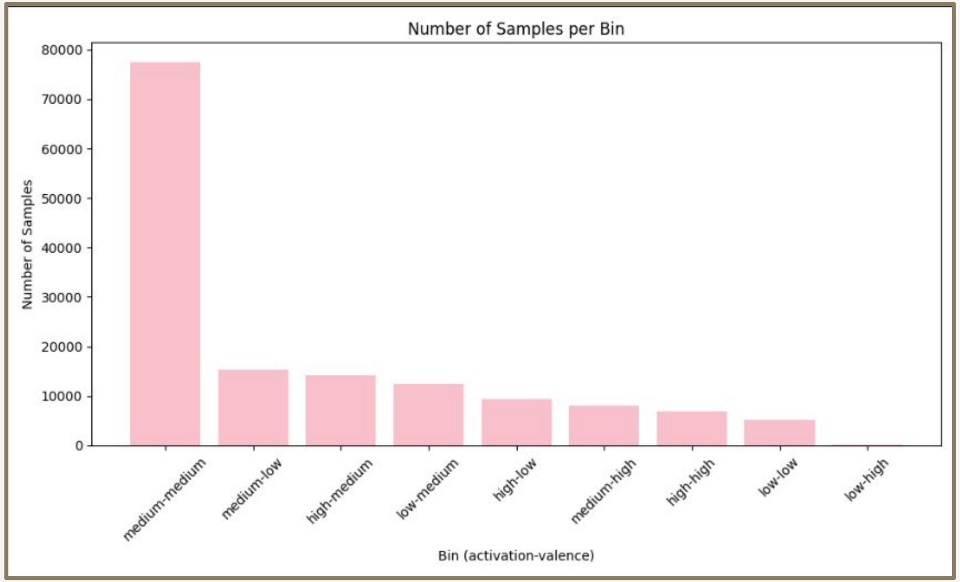
Valence:

- Definition: The intrinsic pleasantness or unpleasantness of an emotion
- Range: Positive (happy) to negative (sad, fear)
- Examples: Feeling good versus bad about a situation



Dr. Megan Anna Neff, *Arousal-Valence Model,* https://neurodivergentinsights.com/arousal-valence-model/?srsltid=AfmBOopIC4OWd5fLnXhwCaa-h-JuZYRDFvfeq5dkvSduSjzn15qy1nla

Activation & Valence Bins



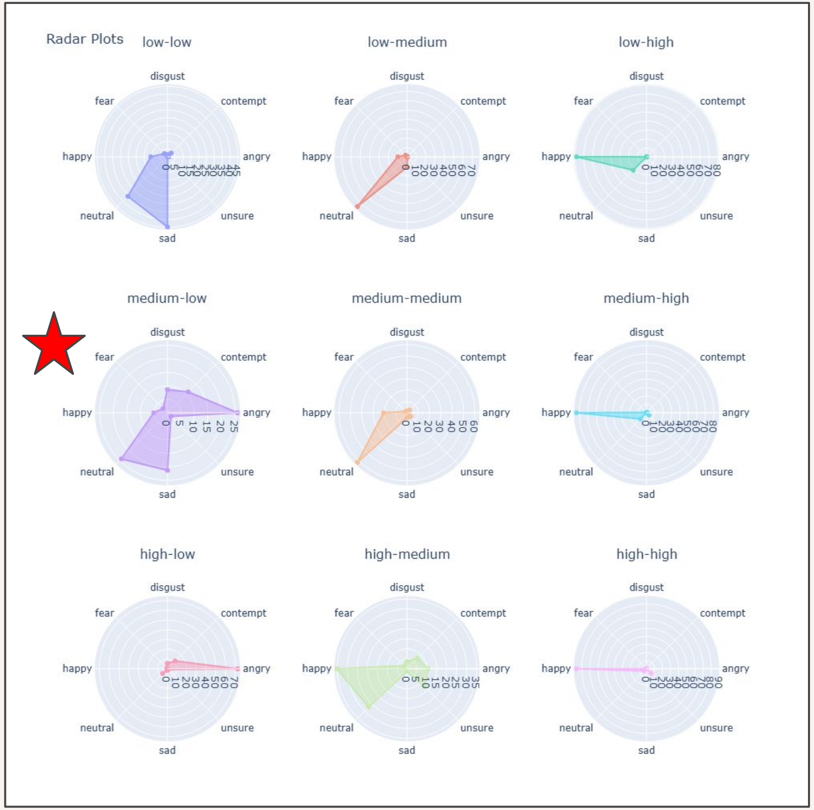
Ratings (1–7) are mapped via fixed cut points at 3 and 5:
Low ≤3, Medium 3 < x ≤ 5, High >5

Number of Samples/bin
(based on CSV output):

Medium-Medium: 77566
Medium-Low: 15328
High-Medium: 14248
Low-Medium: 12522
High-Low: 9298
Medium-High: 8109
High-High: 6849
Low-Low: 5153
Low-High: 234

Total Samples (for entire MSP-Podcast): 149,307

Activation & Valence Radar Plot



Unexpected Outcomes:

- **Medium-Low:** Unexpectedly, the medium–low bin contains non-trivial 'happy' counts (within-bin happy = 3.6% vs overall happy = 19.4%). Although rare, this counters the notion that happy emotions imply high valence. This suggests bin-edge effects or labeling/scale issues. Alternatively, sarcasm/nervous laughter may sound happy while content is negative.

Potential Issues: Neutral dominates across all activation–valence bins (Neutral/bin = 3.7 - 62.9%). This likely reflects annotator uncertainty or class imbalance, so the model learns "neutral" as a safe fallback. For assistive tech, frequent neutral predictions are low-actionable (they don't tell the system what to do next).

03

Ambiguous Samples

High Disagreement Samples

High-disagreement files (for Dev+Test1 dataset): 1,134 / 30,647 (3.70%)
A-only: 436, **V-only:** 643, **Both:** 55

Sample Label (10 randomly sampled (fixed seed)	Disagreement Type	Comments
MSP-PODCAST_1216_0204_0006.wav	**Valence**	Serious topic, excited/disbelief tone
MSP-PODCAST_1657_0079_0002.wav	**Both**	Self-critique, energetic
MSP-PODCAST_1659_0027_0001.wav	**Valence**	Announcer hype vs anger
MSP-PODCAST_0003_0145.wav	**Activation**	Calm voice, excited/gossipy
MSP-PODCAST_0003_0361.wav	**Valence**	Sarcasm/mimicry confounds valence
MSP-PODCAST_0003_0461.wav	**Valence**	Frustration level unclear
MSP-PODCAST_0281_0219.wav	**Both**	Laughing + “oh no” conflict
MSP-PODCAST_0308_0979.wav	**Activation**	Clear laughter, high energy
MSP-PODCAST_0317_0154.wav	**Activation**	Low-energy voice, “interesting” content
MSP-PODCAST_0563_0188.wav	**Valence**	Pleasant tone, negative message

Disagreement metric: Flag if $A\_std \geq \mu+2\sigma$ or $V\_std \geq \mu+2\sigma$.

Analysis

As observed by the number of high disagreement samples for valence versus activation, the skew towards valence indicates that valence is harder to interpret for raters. The sampled clips illustrate why:

- Prosody-Semantic clashes where negative content was conveyed with smiley/pleasant tone
- Sarcasm, mimicry, and laughter that resemble “happy” when the context does not

In contrast, activation disagreements tend to stem from atypical energy cues:

- Calm delivery of supposedly interesting stories
- Laughter/excited speech where non-lexical events misconvey perceived arousal

Cases flagged on both properties usually combine conflicting signals or rapid emotional shifts within the segment. Practically, these regions are likely failure modes for SER and help explain the neutral bias. When annotators diverge, models also default to neutral, which lowers downstream actionability.