

# Building an Emotionally Aware Dialogue System with Memory

Vanshika Reddy

Senior Thesis, CSSE Department, Rose-Hulman Institute of Technology

## Abstract

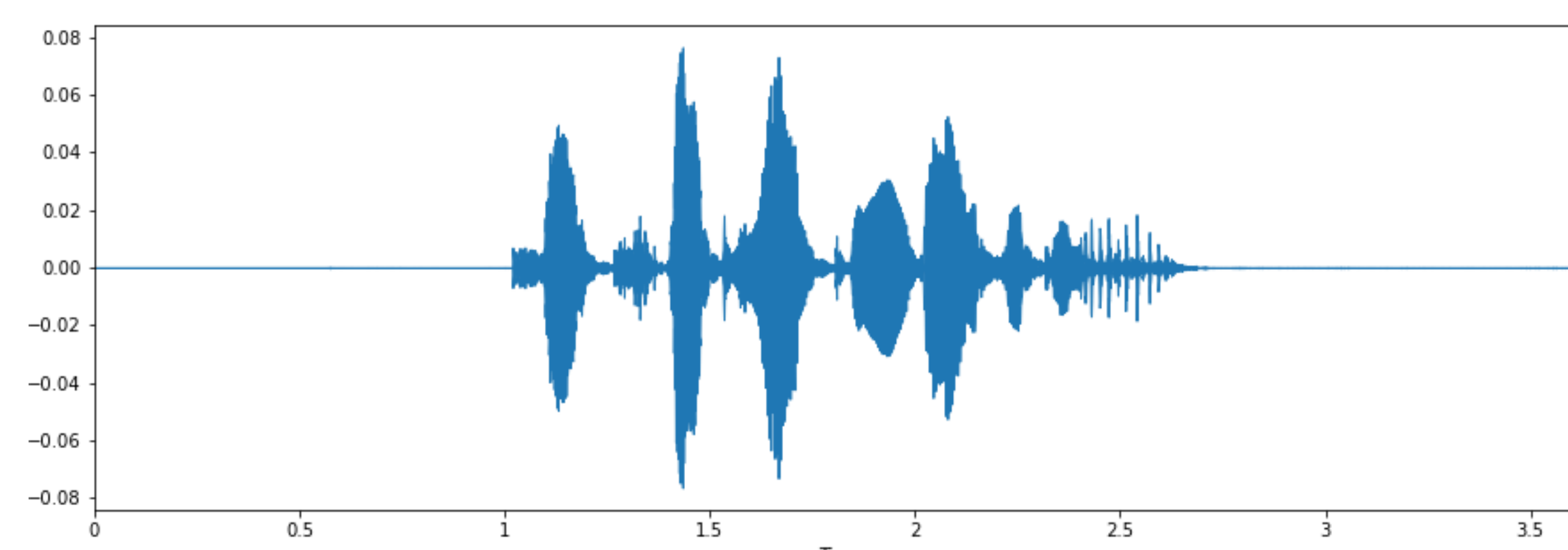
This project proposes a unique way to build a memory along with analyzing the user's input to generate emotionally appropriate responses. The model contains two important and distinct features: generating a mental model, which acts as memory for the system and analyzing the intensity of the emotion. By remembering instances and intensities of an emotional event, the model tries to etch out an emotional profile of the user which is a key input in the response generation process. The contribution of this project is a system to better understand human emotions and provide human-like emotional assistance

## Objectives

- Build a system that can detect human emotions and generate individually tailored responses
- Build a memory system that stores all important information and models the users' emotional responses to incidents

## Data

The dataset includes around 1500 audio file inputs from 24 different actors capturing 8 emotions that can be detected by the emotional model. The data contains short recordings of 12 male and 12 female actors to ensure a balance. The data was split into an 80:20 training : test split and the training data was further broken down into training and validation sets.



## Software and Technology

- Python
- Deep Learning
- TensorFlow
- Opensource voice frequency detection model (MIT Licensed)

## Architecture

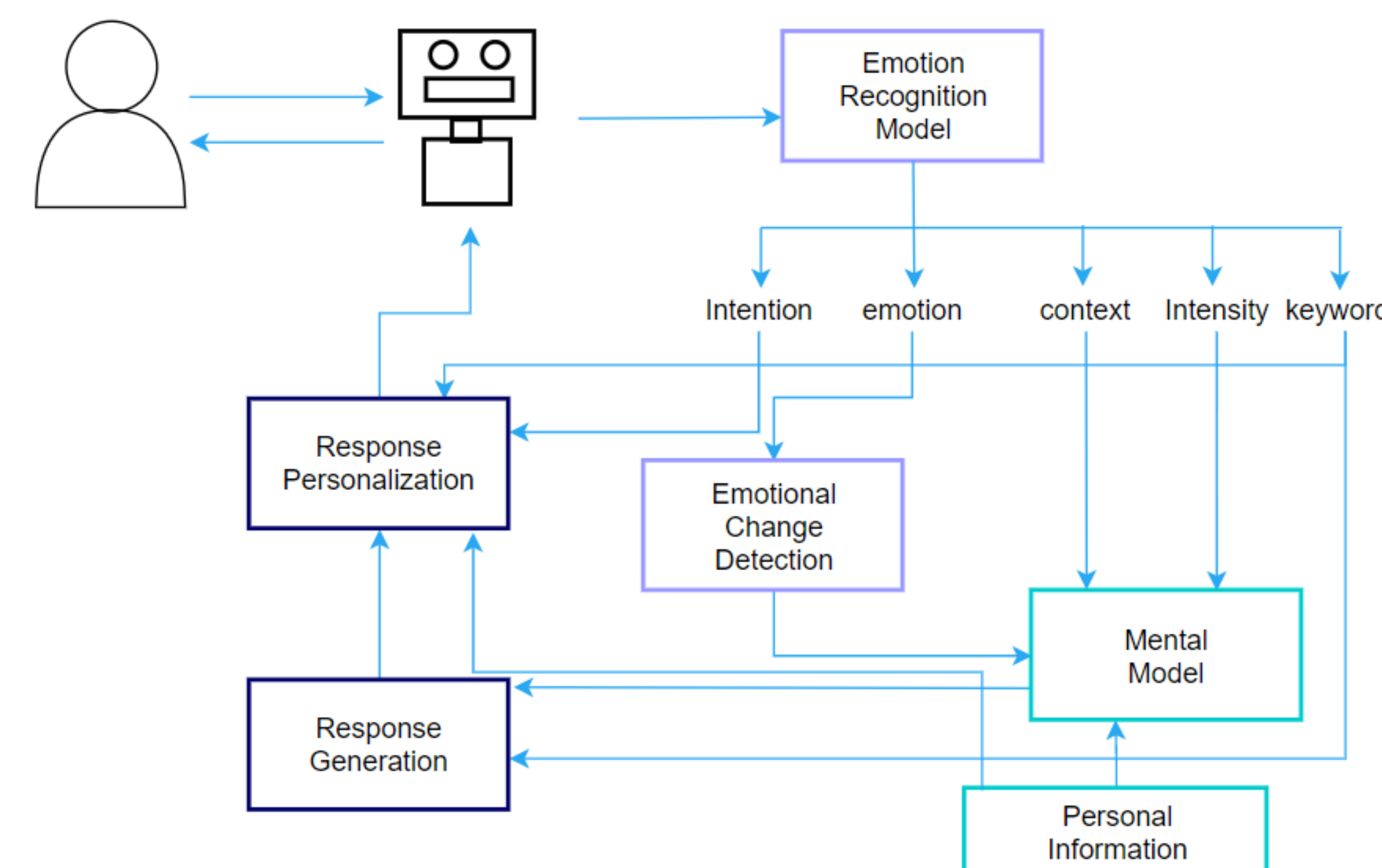
The System contains 3 major components:

- **Emotion Recognition Model**
- **Mental Model**
- **Response generation**

This Thesis Project only focuses on the first 2 components and is currently only processing one emotion (sad).

The System can be expanded to other emotions without changing the system only by modifying and retraining the emotional model.

So far, the Emotion Recognition model and the Mental Model are the parts that have been explored/implemented.



## Mental Model

- Acts as a whole data storage unit
- Analyzes the user's sentences
- Stores all the necessary information

The data is stored with the following crucial information:

- **Keywords**
- **Time**
- **Emotional state**
- **Change in emotion and cause of change**
- **Response to be given / customizable responses**

It is a Mapped structure with keywords mapped to linked lists of instances which could be interlinked in the mapping.

## Emotion Recognition Model

The Model uses 2 different ways of classifiers to individually classify emotions and intensity.

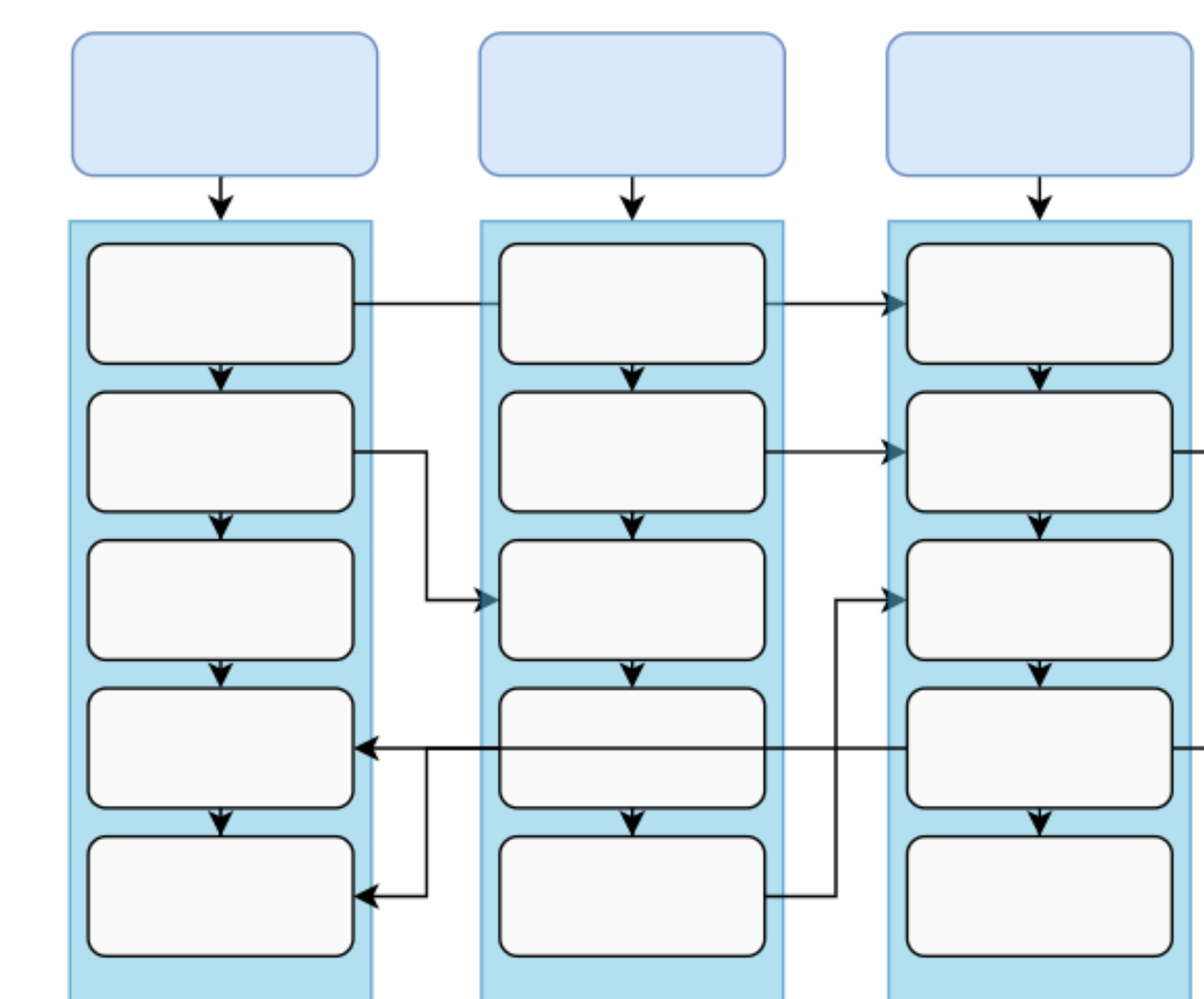
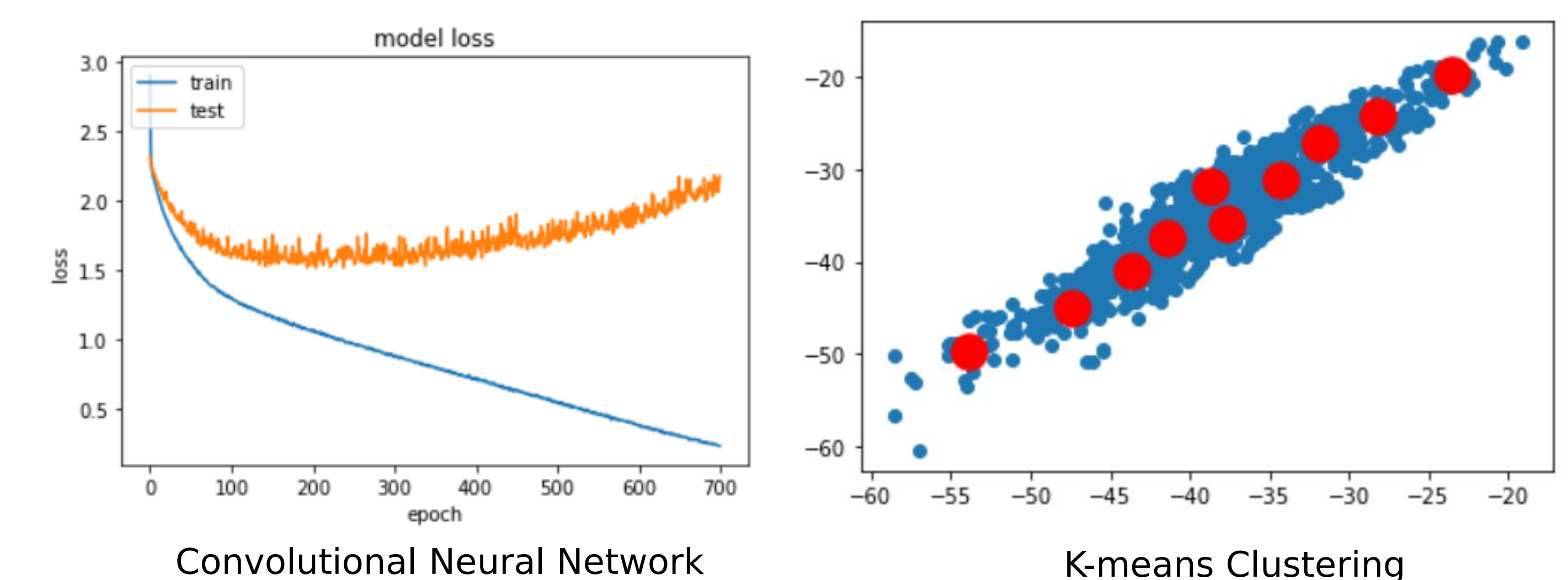
- To classify emotions the Model uses a Convolutional Neural Network to detect between 8 emotional states
- 1 = neutral, 2 = calm, 3 = happy, 4 = sad, 5 = angry, 6 = fearful, 7 = disgust, 8 = surprised
- The data then also goes through a K-means clustering algorithm with k=10 to find 10 different intensities
- The Intensity is based on using the average frequency and median frequency as features on the 3 second recording.

The CNN was able to achieve an accuracy of **73%** on its best run on the test set.

It is extremely hard to figure out the accuracy of the k-means clustering algorithm because the intensity opinions would be based on my ratings and would be biased.

Data needs to be cleaned and classified to get better results on Intensity classification

The emotion and the intensity together form a three-digit integer representing the emotional state of the person where the first digit represents the emotion, and the last two-digits represent the intensity



Mental Model