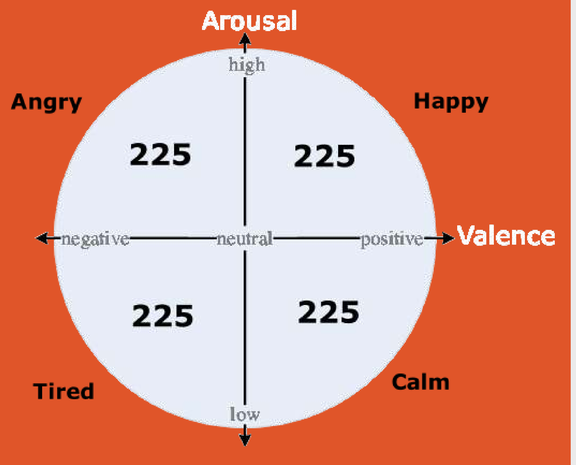
MUSIC SENTIMENT ANALYSIS

Predicting emotion in music using deep learning

EMOTION PREDICTION PROVES CHALLENGING

- Emotion classification of music has the potential to provide benefits across many fields including personalized music recommendations, mental health applications, and music composition.
- However, this task has proved to be challenging. This is evidenced by competitions held at conferences like MediaEval and Music Information Retrieval Evaluation eXchange (MIREX) which have yet to solve the issue.
- Renato Panda, a professor and researcher at the University of Coimbra achieved the best accuracy results at MIREX in 2012. Currently, the results of state-of-the-art models in this field have stagnated and have yet to significantly pass his results.
- 4Q dataset: The dataset we use was created by Panda to address this stagnation issue in a dataset which consists of 900 songs. These songs are discretely labeled in the psychologically based four quadrants of James Russel's Valence Arousal Circumplex Model.

Figure 1: Valence/Arousal model with the distribution of songs in the 4Q dataset.



Oregon State University

REPLICATION OF *NOVEL AUDIO FEATURES FOR MER* - PANDA

In this replication study of Renato Panda's work, we successfully achieved comparable results using the same 4Q dataset and same extracted features.

- From the 71 musical dimensions in figure 4, static statistical features were extracted from the audio signal e.g., mean of the power spectrum.
- From a total of >1700 statistical features, the top 100 were chosen using feature selection.
- These 100 features were evaluated on a Support Vector Machine (SVM) to classify emotion.
- Like Panda, we used 10-fold cross validation across 20 runs to verify repeatability.
- Achieved 76.8% F_1 (figure 2) vs Panda's 76.4%.
- In addition to achieving a similar F_1 as Panda, we also found that quadrants three and four are the hardest to classify as seen in Figure 2.

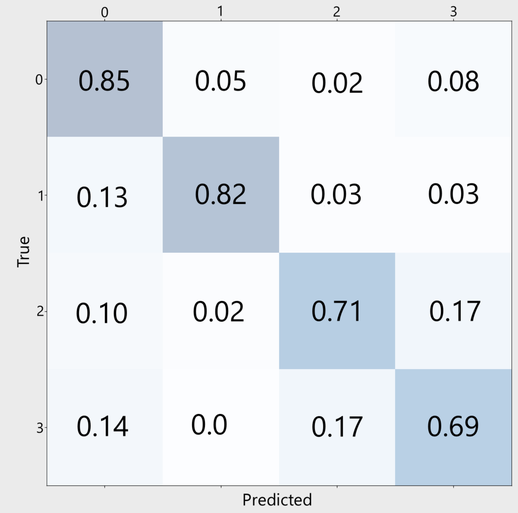


Figure 2: Confusion matrix showing classification using static features on SVM

EMOTION CLASSIFICATION USING RAW AUDIO & SPECTROGRAMS

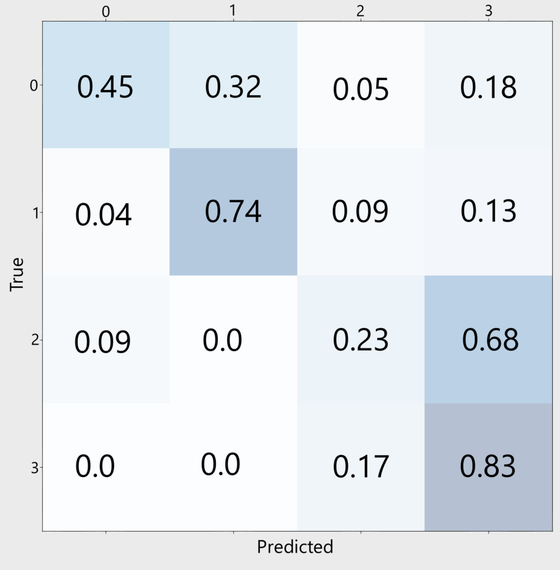


Figure 3: Confusion matrix showing classification using spectrograms on CNN

We used deep learning to classify emotion using two different modalities to improve classification accuracy. These modalities were chosen to avoid the time-consuming preprocessing step of feature extraction.

Raw Audio

- Instead of extracting features, we used the raw audio file (.wav) as input to our deep learning model.
- We evaluated this raw audio using a Bidirectional Long-Short Term Memory (BLSTM) model.
- ~ 33% F_1

Spectrograms

- Spectrograms are visual representations of audio signals over time at various frequencies.
- We evaluated these spectrograms using a Convolutional Neural Network (CNN).
- ~ 55% F_1 (Figure 3)

Feature Type	Number Features
Rhythm	10
Dynamics	9
Melody	5
Harmony	12
Tonal Color	28
Musical Form	2
Expressive Tech.	1
Other	4
Total	71

Figure 4: Table showing the distribution of features per musical dimension

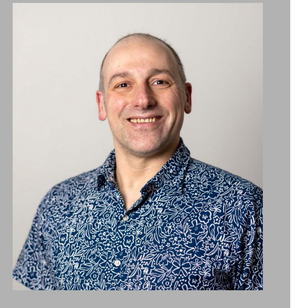
CONCLUSION & NEXT STEPS: DYNAMIC FEATURE EXTRACTION

- Our results using raw audio & spectrograms were not quite as good as extracted features even when using deep learning. This shows that this is a much harder problem & more robust methods need to be researched.
- In the future, we plan on extracting dynamic features for each audio file instead of static ones.
- This will provide us data that dynamically changes and better represents the audio.
- These features can be the input for a future BLSTM, or part of a larger multi-modal model which can accurately classify all four quadrants.

OUR TEAM

- Shaun Leib (Left)
- Mentor: Professor Patrick Donnelly (Right)





soundbendor lab