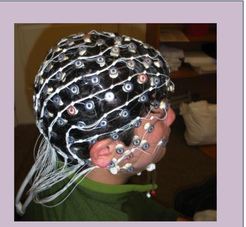
04/26/2024

Music Tempo Estimation From EEG Data

UAA College of Engineering
University of Alaska Anchorage

Eleonora Stadtmüller Caballero [1], ***Jaren Ramirez*** [1], ***Quinton Odenthal*** [1]
with faculty advisors ***Dr. Blair Kaneshiro*** [1,2] *and* ***Dr. Masoumeh Heidari*** [1]

[1] Department of Computer Science and Engineering, University of Alaska Anchorage, [2] Center for Computer Research in Music and Acoustics, Stanford University



Introduction

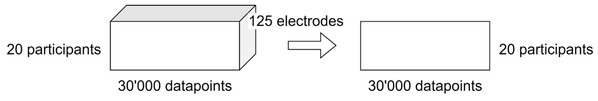
Estimating the tempo of a song is considered one of the most fundamental tasks of music information retrieval (MIR), but it is mostly performed using audio waves. For our project, we explore the feasibility of estimating a song's tempo from electroencephalogram (EEG) data collected from individuals listening to the song using multiple machine learning techniques.

Dataset

- **Name:** Naturalistic Music Estimation Dataset - Tempo (NMED-T)
- **Data Type:** EEG time-series data stored in 3D Matlab matrices
- **Size:** 200 samples, 20 individuals x 10 songs
- **Tempo Range:** 55.97 bpm - 150 bpm
- **Challenges:**
  - EEG data is very noisy
  - Small dataset
  - Octave error

Data Preparation

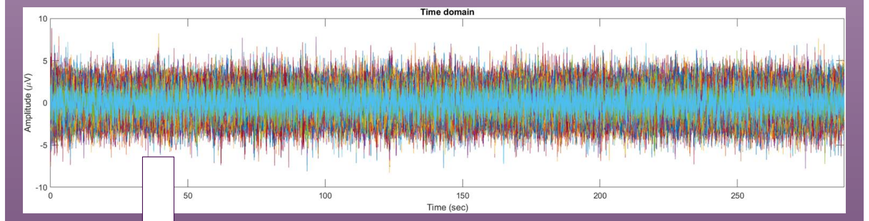
- **Reliable Component Analysis** (RCA) was used to reduce the dimensionality of the data along the **electrode dimension** per song.



- Data was standard scaled and normalized using **l2 normalization**.
- For multiclass classification, **labels** were added to each song based on a tempo bucket and classes were **weighted** based on their distribution.
- Data was **augmented** by splitting the song data into smaller chunks to **counteract overfitting** due to the small size of the dataset.

Methods and Models

- **K-fold cross validation** was used to generate more **reliable** performance estimates.
- **1D CNN** (1-dimensional Convolutional Neural Network) for **multiclass classification.**
- 2D CNN (2-dimensional Convolutional Neural Network) for **regression**.

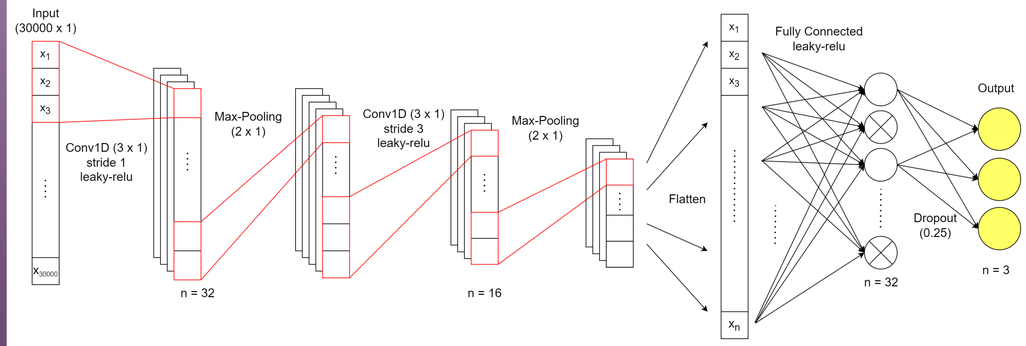


"103.7 bpm"

We can **estimate** a **song's tempo** from the **brain waves** of someone who **listened** to that song.

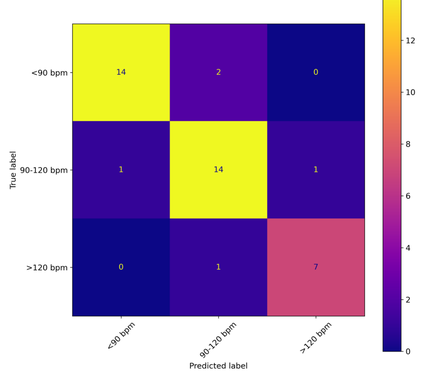
Model Architecture

- 1D CNN for multiclass classification (3 classes)

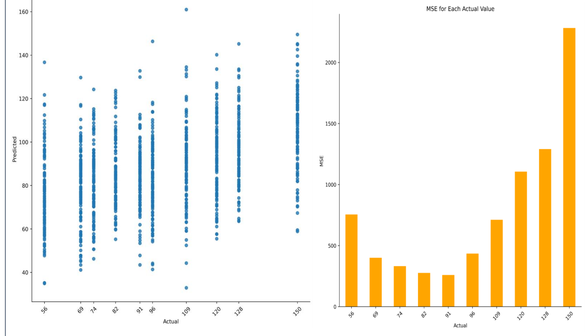


Results

- 1D CNN for multiclass classification
  - 3 classes (below): Accuracy 87.5%
  - 5 classes: Accuracy 85%



- 2D CNN for regression
  - predicted vs actual (below)
  - Error increases with higher bpm



Conclusions and Future Work

- Both the **multiclass classification and regression** models generally performed **better** when evaluated on songs in the **high density medium tempo range** (~90 bpm) and **worse** for songs in the **slow** (~55 bpm) or **fast** (~150 bpm) dataset tempo extremes.
- Making the model **too deep or complex** led to **faster overfitting** and **worse performance**.
- To explore in future work:
  - Fusing both the **time and frequency domains** of the EEG data into a **two-feature model** could improve performance.
  - For the regression model, accounting for the **higher density** of data in the **medium tempo range** using a weighting technique could improve performance.
  - Making the **loss invariant to the octave error** experienced by different subjects could provide a **more correct estimation** of the tempo.