Carnegie Mellon University

Deep Music Features for Music Recommendation Systems Proposal

Nicholas Magal

Problem Statement

Recommender Systems

Recommender Systems
Content Based Filtering vs Collaborative Filtering

Content Based Filtering

Recommend items based on similar liked items

Advantages

- No data needed from other users
- Can recommend niche items that other users are not interested in

Disadvantages

- Need to hand craft features
- Limited Recommendations

Collaborative Filtering

Recommend items based on similarities between users and liked items

Advantages

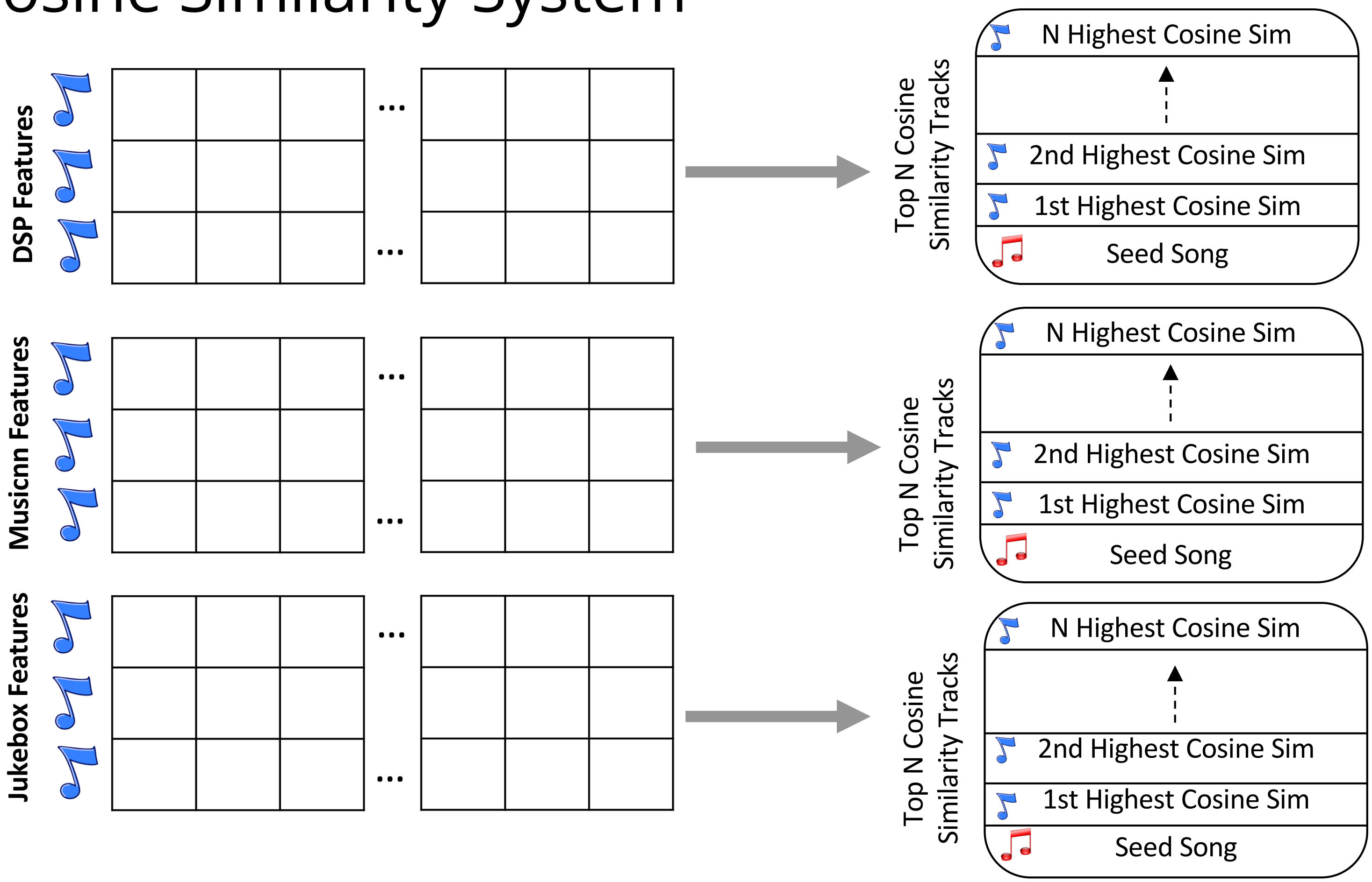
- Incorporates more information
- No domain knowledge needed

Disadvantages

- Cold start problem

Carnegie Mellon University

Proposed Approach
Cosine Similarity System



Carnegie Mellon University

Dataset

GTZAN

- 1,000 half minute music excerpts
- 10 different genre labels
  - Blues, classical, country, disco, hiphop, jazz, metal, pop, reggae, rock

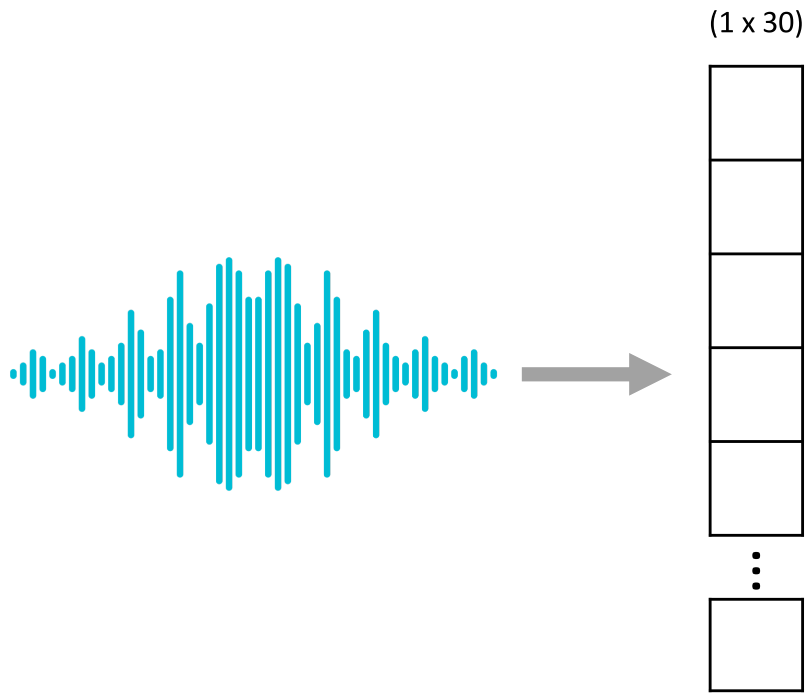
Carnegie Mellon University

Citations: [3]

Feature Representations

DSP Features
Timbral Texture Features

- Mean and variance of:
    - Spectral Centroid
        - Brightness
    - Spectral roll-off
        - Frequency concentration
    - Zero crossing rate
        - Noisiness
    - 12 MFCCs

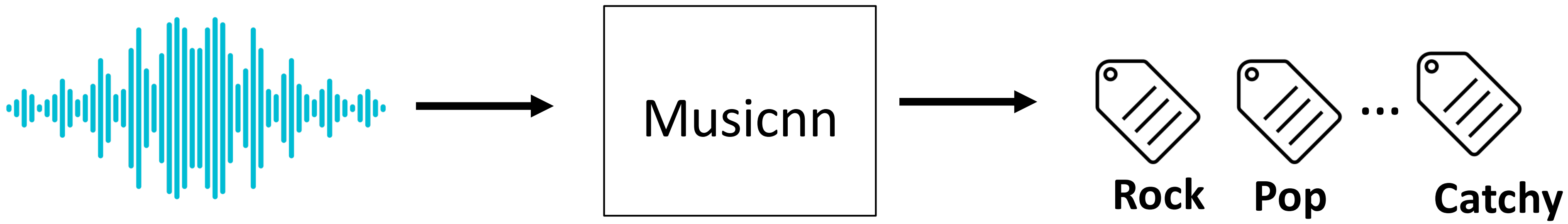


Carnegie Mellon University

Deep Learning Features

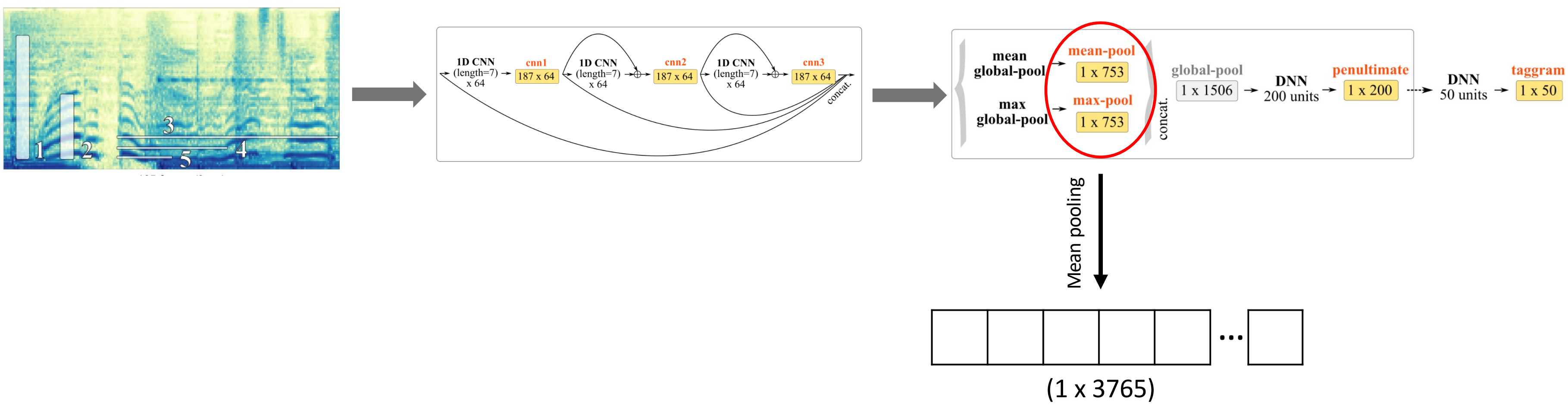
Musicnn Features
High Level Overview

- Musically motivated convolutional neural network
- Pretrained on Million Song Dataset for auto-tagging



Carnegie Mellon University

Musicnn Features
Architecture



Carnegie Mellon University

Citations: [5]

Jukebox Features

High Level Overview

- Generative Model based on Hierarchal VQ-VAE and Transformer architecture
- Pretrained on 1.2 million songs scraped by OpenAi
- Performs 30% on average better then other embeddings on genre detection, music emotion classification, auto-tagging, and key detection

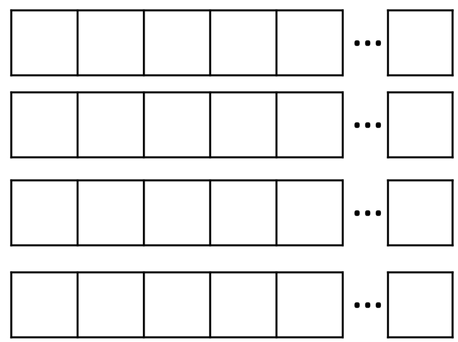
Carnegie Mellon University

Citations: [6]

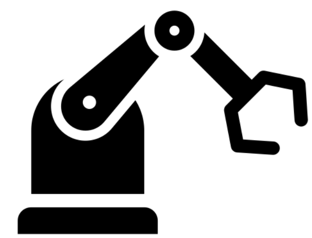
Jukebox Features

High Level Overview

1.) Audio Input



2.) Codified Discrete Representations

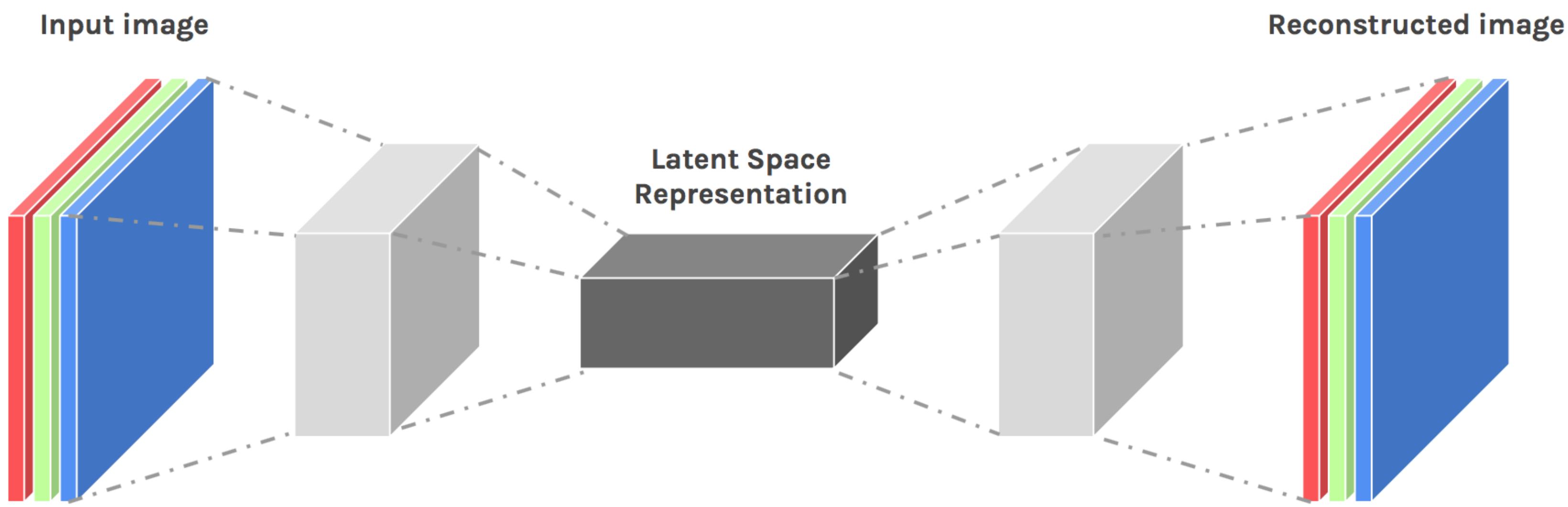


3.) Model

Citations: [6]

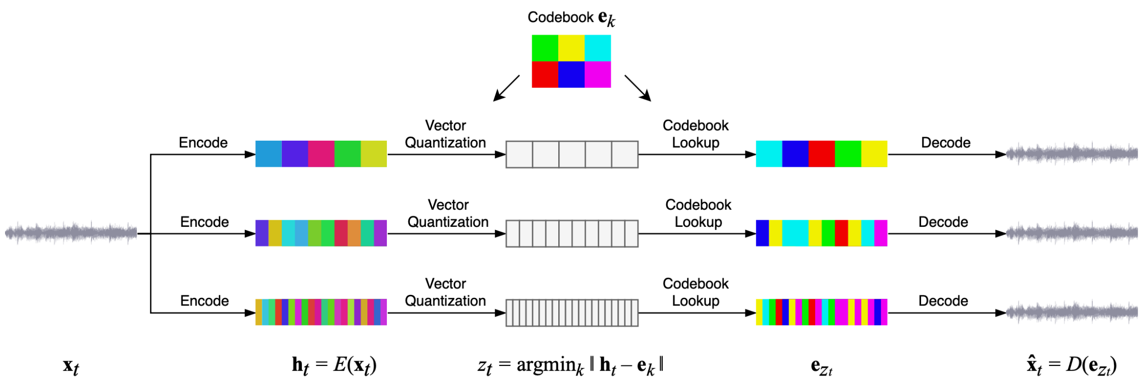
Quick Detour

Autoencoders



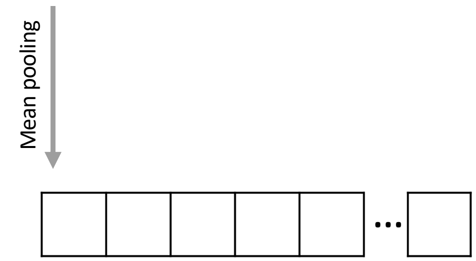
Jukebox Features
Model Details

1.) Generate Codified Discrete Representations



2.) Model Codified Discrete Representations

$$p(\mathbf{z}) = p(\mathbf{z}^{\text{top}}, \mathbf{z}^{\text{middle}}, \mathbf{z}^{\text{bottom}})$$



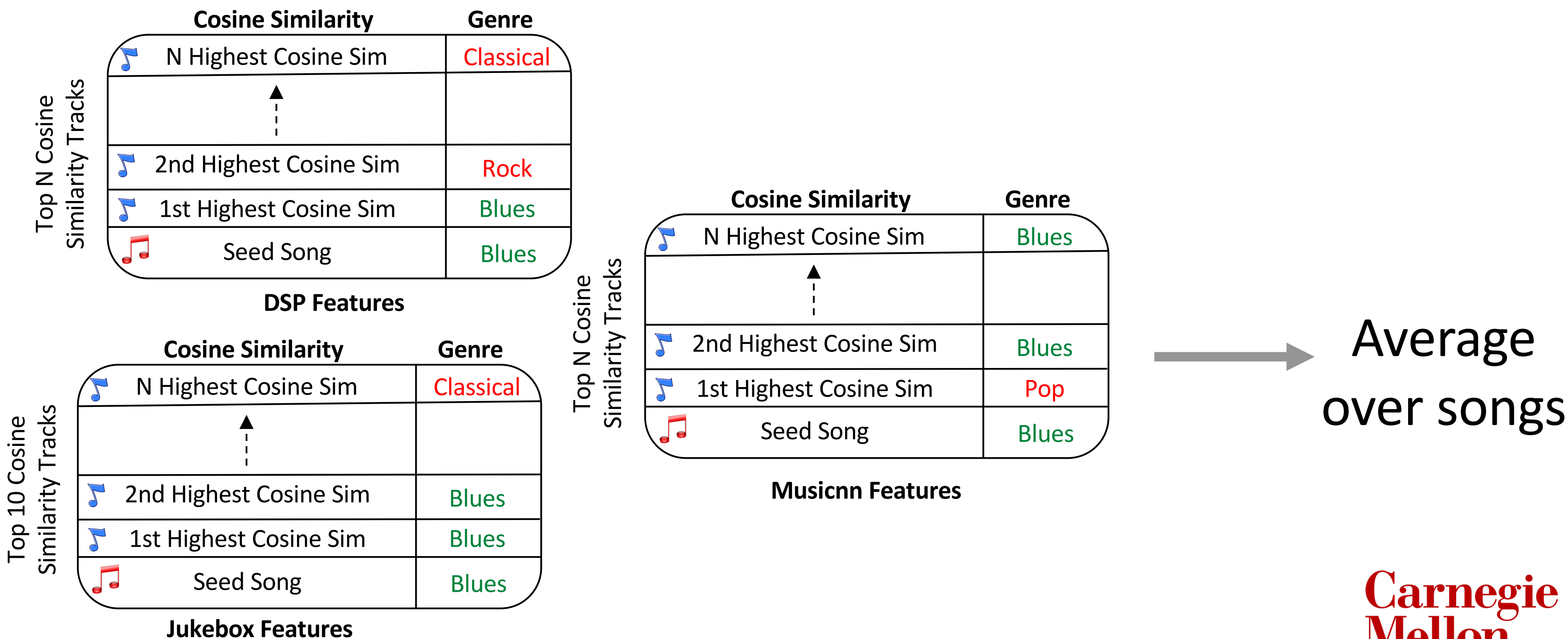
(1 x 4800)

Citations: [6]

Results

Quantitative Evaluation

Average Matching Genre of Top N Tracks



Carnegie Mellon University

Quantitative Evaluation

Average Matching Genre of Top 10 Tracks

Top 10 Cosine Similarity Tracks

Cosine Similarity	Genre
10 Highest Cosine Sim	Classical
⋮	
2nd Highest Cosine Sim	Rock
1st Highest Cosine Sim	Blues
Seed Song	Blues

DSP Features

Top 10 Cosine Similarity Tracks

Cosine Similarity	Genre
10 Highest Cosine Sim	Classical
⋮	
2nd Highest Cosine Sim	Blues
1st Highest Cosine Sim	Blues
Seed Song	Blues

Jukebox Features

Top 10 Cosine Similarity Tracks

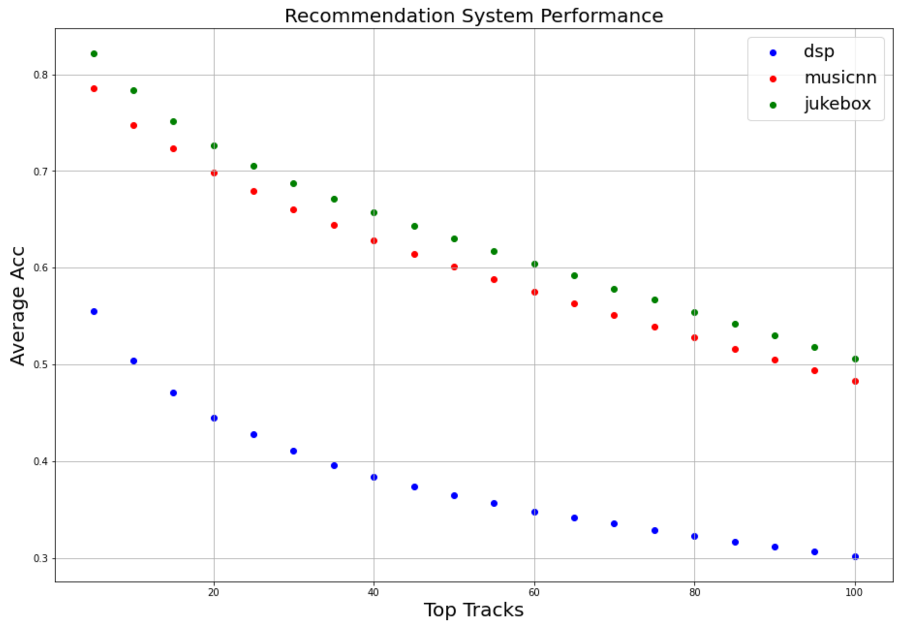
Cosine Similarity	Genre
10 Highest Cosine Sim	Blues
⋮	
2nd Highest Cosine Sim	Blues
1st Highest Cosine Sim	Pop
Seed Song	Blues

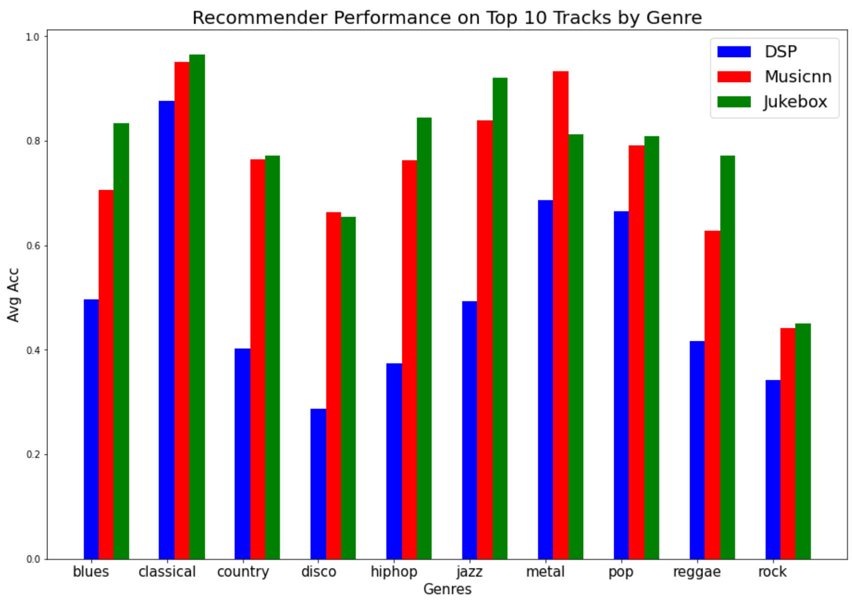
Musicnn Features

→ Average over genres

Carnegie Mellon University

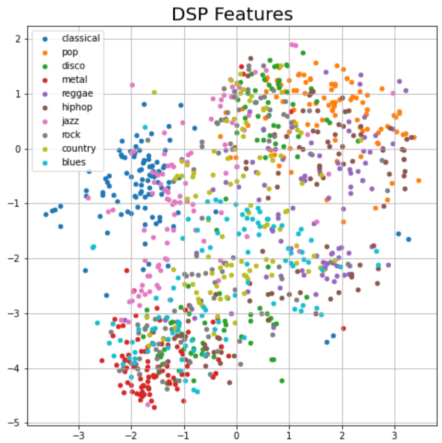
Results

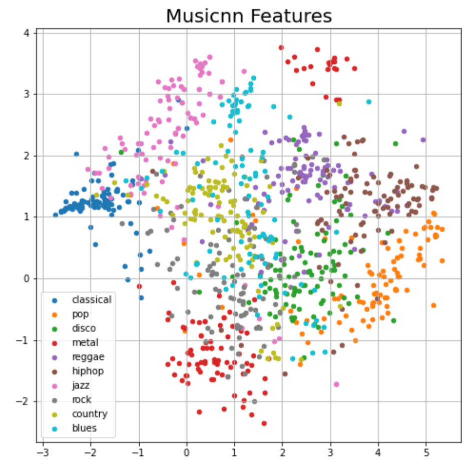


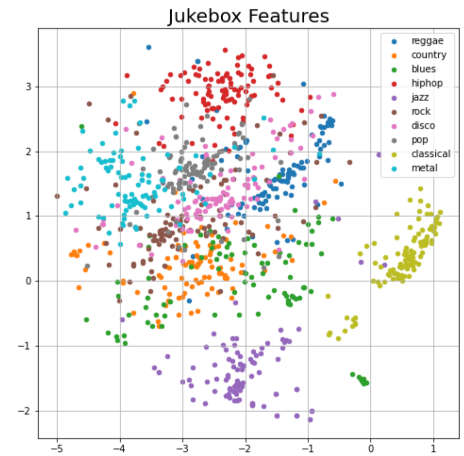


Carnegie Mellon University

Results
TSNE Plots







Carnegie Mellon University

Proposed Auditory Evaluation

Online Survey:

1. 3 prechosen seed songs
2. From each feature set, return the top N most similar song
3. User rates the recommendation on a scale of 1-10 based on lyrics, rhythm, melody, and overall traits of each track.

Thank You!

Citations

[1]: Google Developers, “Content-based filtering.”
`https://developers.google.com/machinelearning/recommendation/content-based/basics`, Jul 2018

[2]: Google, “Collaborative filtering.”
`https://developers.google.com/machinelearning/recommendation/collaborative/basics`, Jul2018.

[3]: Sturm, Bob L. "An analysis of the GTZAN music genre dataset." *Proceedings of the second international ACM workshop on Music information retrieval with user-centered and multimodal strategies*. 2012.

[4]: G. Tzanetakis and P. Cook, “Musical genre classification of audio signals,” IEEE Transactions on speech and audio processing, vol. 10, no. 5, pp. 293–302, 2002.

[5]: J. Pons and X. Serra, “Musicnn: pre-trained convolutional neural networks for music audio tagging,” 2019.

[6]: Castellon, Rodrigo, Chris Donahue, and Percy Liang. "Codified audio language modeling learns useful representations for music information retrieval." *arXiv preprint arXiv:2107.05677* (2021).