Neural Notes

Visual Exploration of Songs using AI Content-Based Embeddings

W210 Capstone Project, Week 14 Final

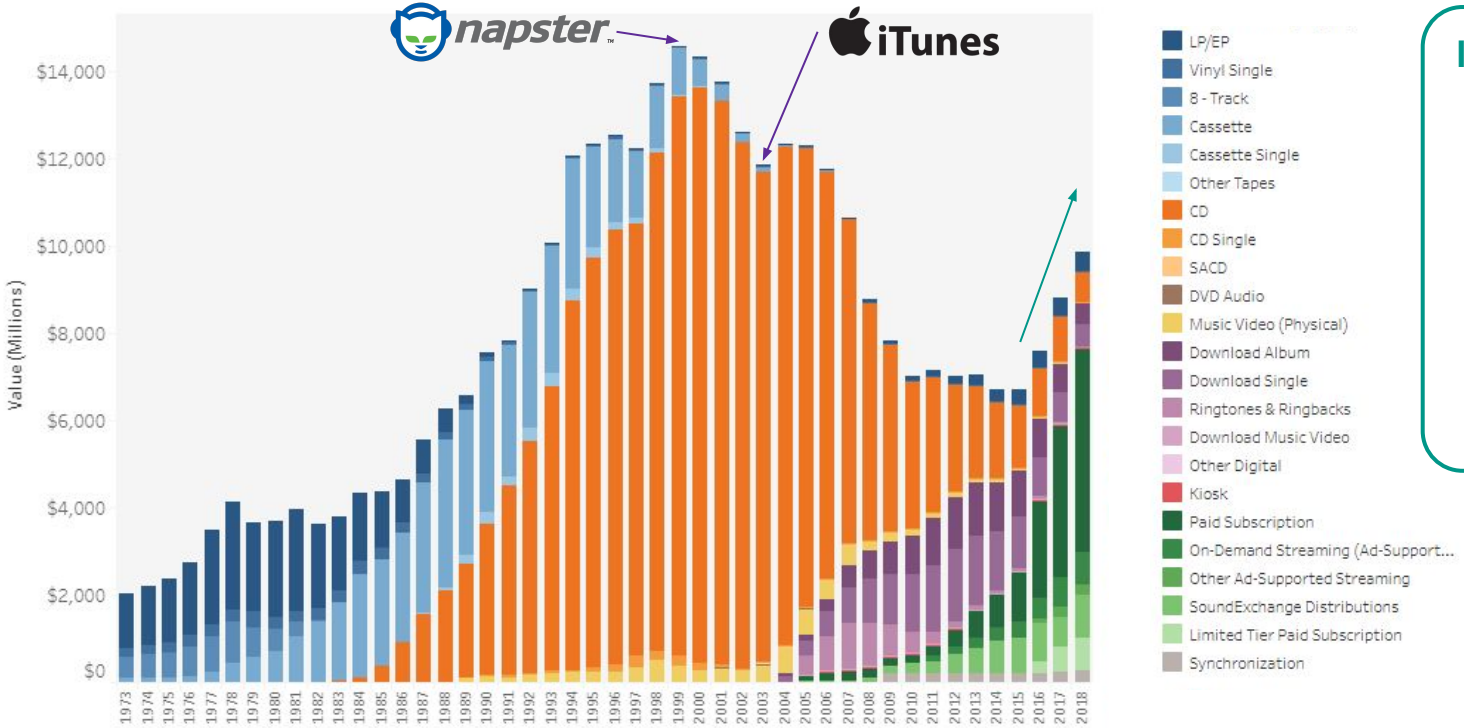
Madeleine Bulkow | Kuangwei Huang | Weixing Sun

Agenda

1. **Introduction**
   - Background of the Music Industry
   - Opportunities in Streaming Music
2. **Problem Statement and Our Solution**
   - Demo: Neural Notes
3. **How we did it**
   - Spectrograms
   - Transfer Learning with CNN
   - Song Embeddings
   - T-SNE Visualization
4. **Neural Notes: MVP Recap**
5. **Conclusions and Future Work**
6. **Group Details**

Background: Music Industry

U.S. Recorded Music Revenues by Format 1973 to 2018



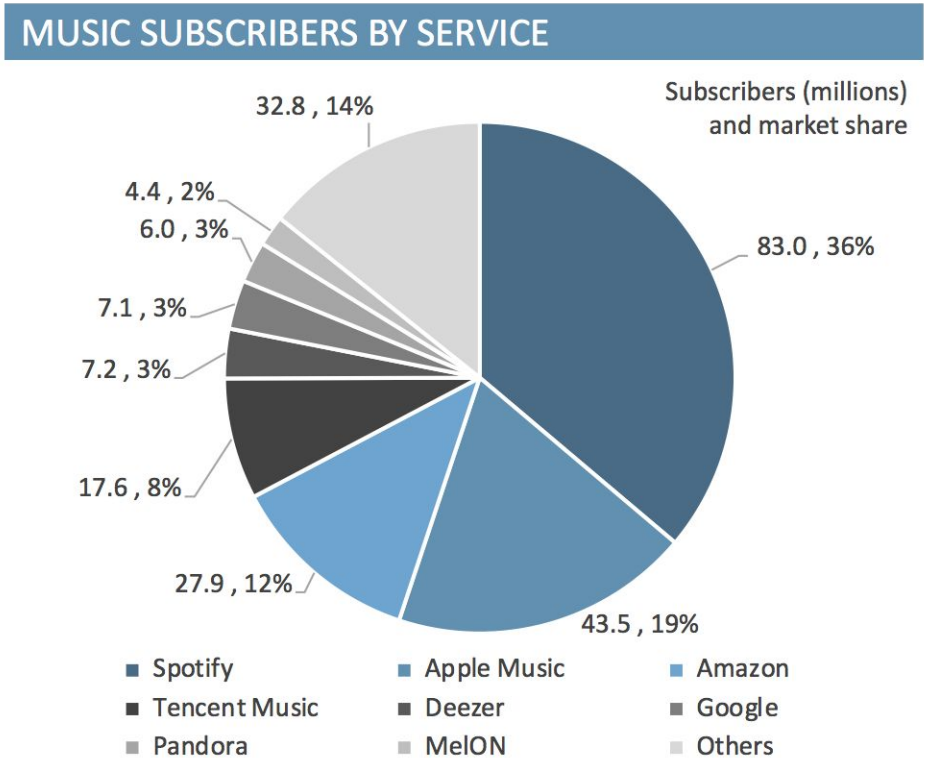
In 2018:

- $9.2 billion
- 12% YoY growth
- 75% streaming
- > 50 mil paid subscriptions

Source: https://www.riaa.com/u-s-sales-database/

Streaming Music: Opportunities

- Globally streaming revenues of $8.9B in 2018
- Streaming revenues grew by 34.0% in 2018
- Platforms compete on personalized content and “discovery”
- Recommender systems, traditionally content-agnostic
- Opportunity for content-based recommendation using deep learning
- Song profiling, akin to NLP word embeddings



Source: https://www.ifpi.org/news/IFPI-GLOBAL-MUSIC-REPORT-2019
Source: https://www.midiaresearch.com/app/uploads/2018/09/midia-mid-year-2018-subscriber-mareket-shares.png

Problem Statement

"How might we have content-based song recommendations, and a new way to "discover" music on streaming platforms?"

Proposed Solution

deep learning techniques

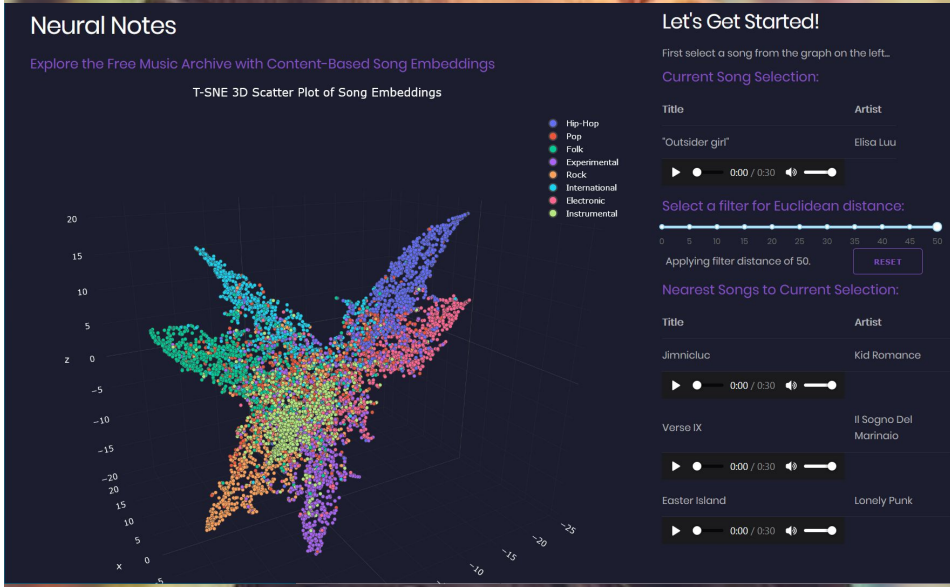
↓

content-based profiling of songs

↓

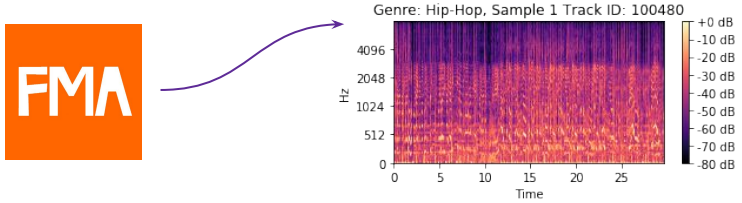
new approach to song recommendation and "discovery" experience

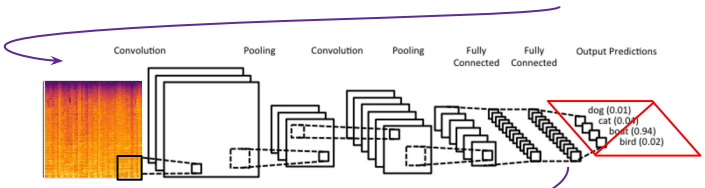
Demo

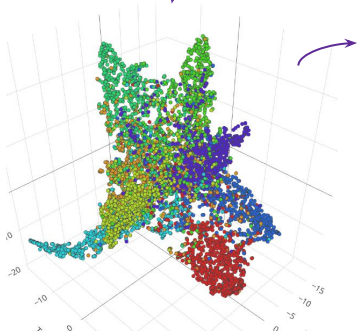


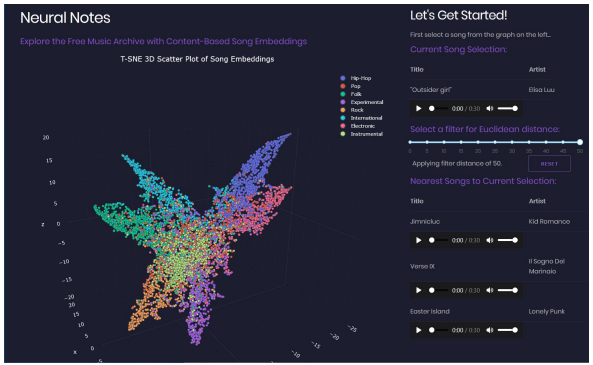
https://neural-notes.appspot.com/

How we did it.









- Dataset: Free Music Archive
- Start with genre classification
- Converting audio samples into spectrograms
- Transfer Learning with CNN
- Song embeddings
- T-SNE visualization of high dimensional embeddings
- New visual “discovery” experience for streaming music

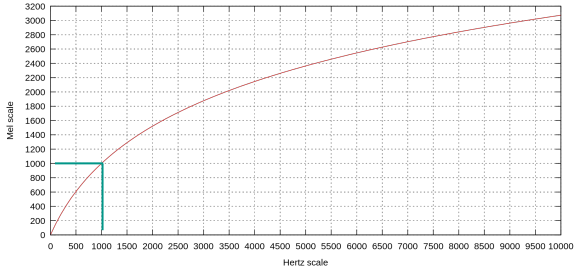
Data Extraction: Spectrograms

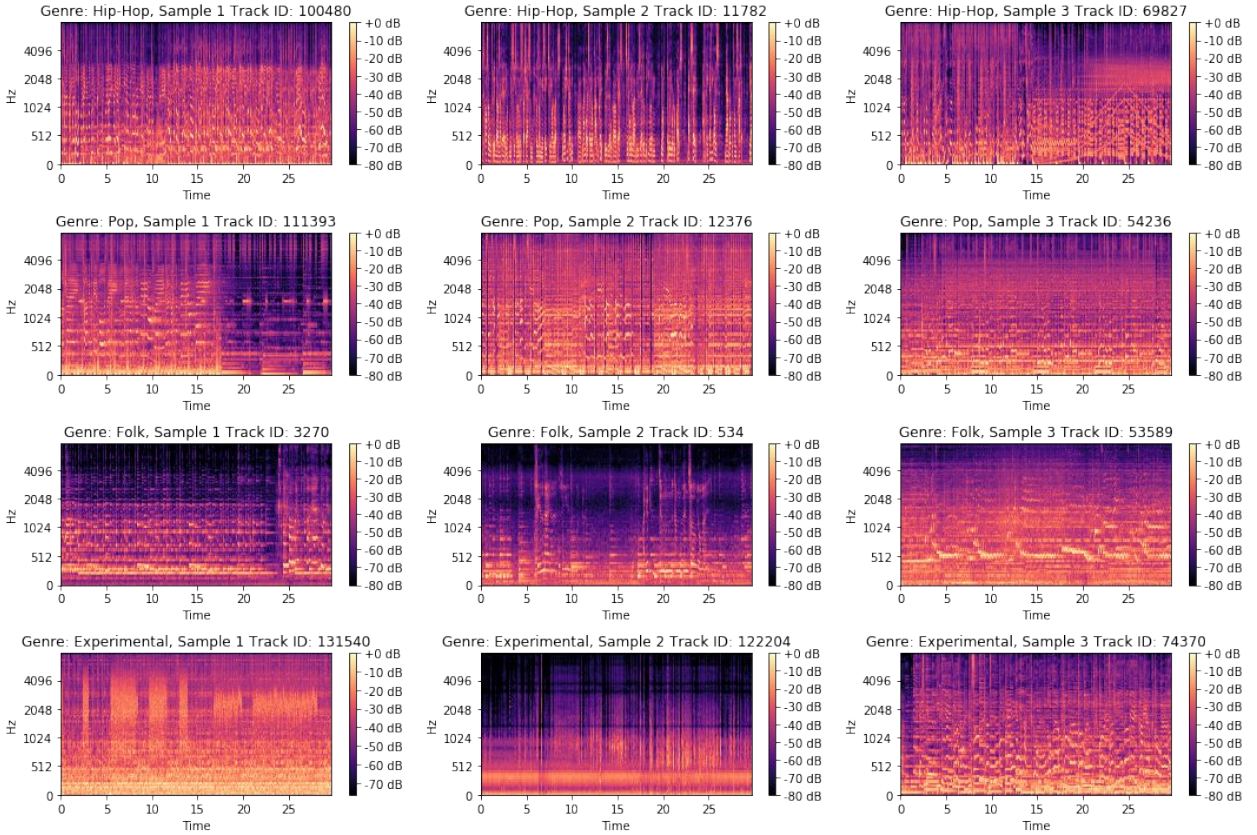
Data: Free Music Archive, small - 8000 samples, balanced 8 genres.

Goal: convert time-series audio signals to spectrograms for image recognition with CNN.

Spectrogram:
Frequency map with decibel intensity over the time duration

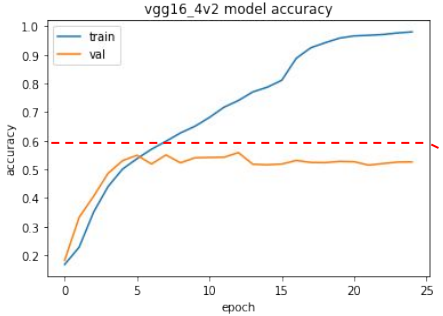
Mel scale:
Psycho-acoustic scale of pitch

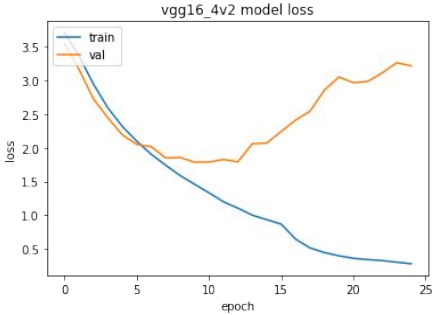




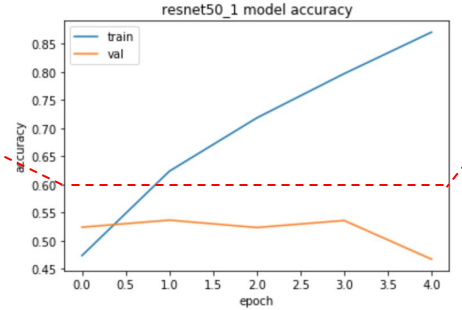
Transfer Learning with Pre-trained CNNs

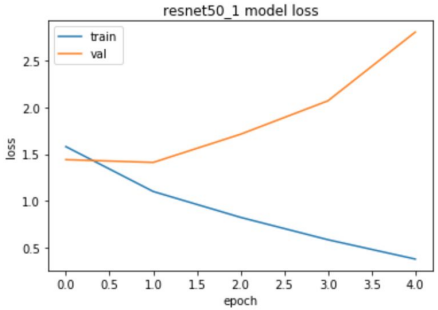
VGG16
Accuracy: 0.56



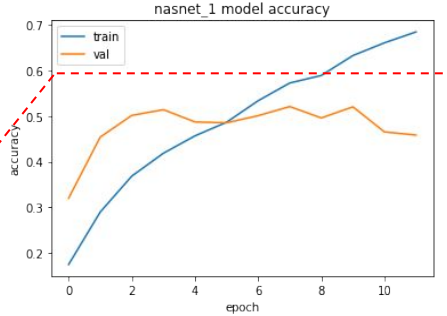


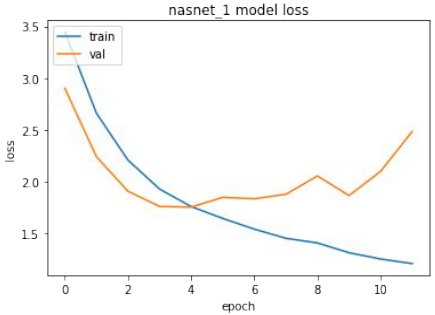
ResNet
Accuracy: 0.53





NASNet
Accuracy: 0.52





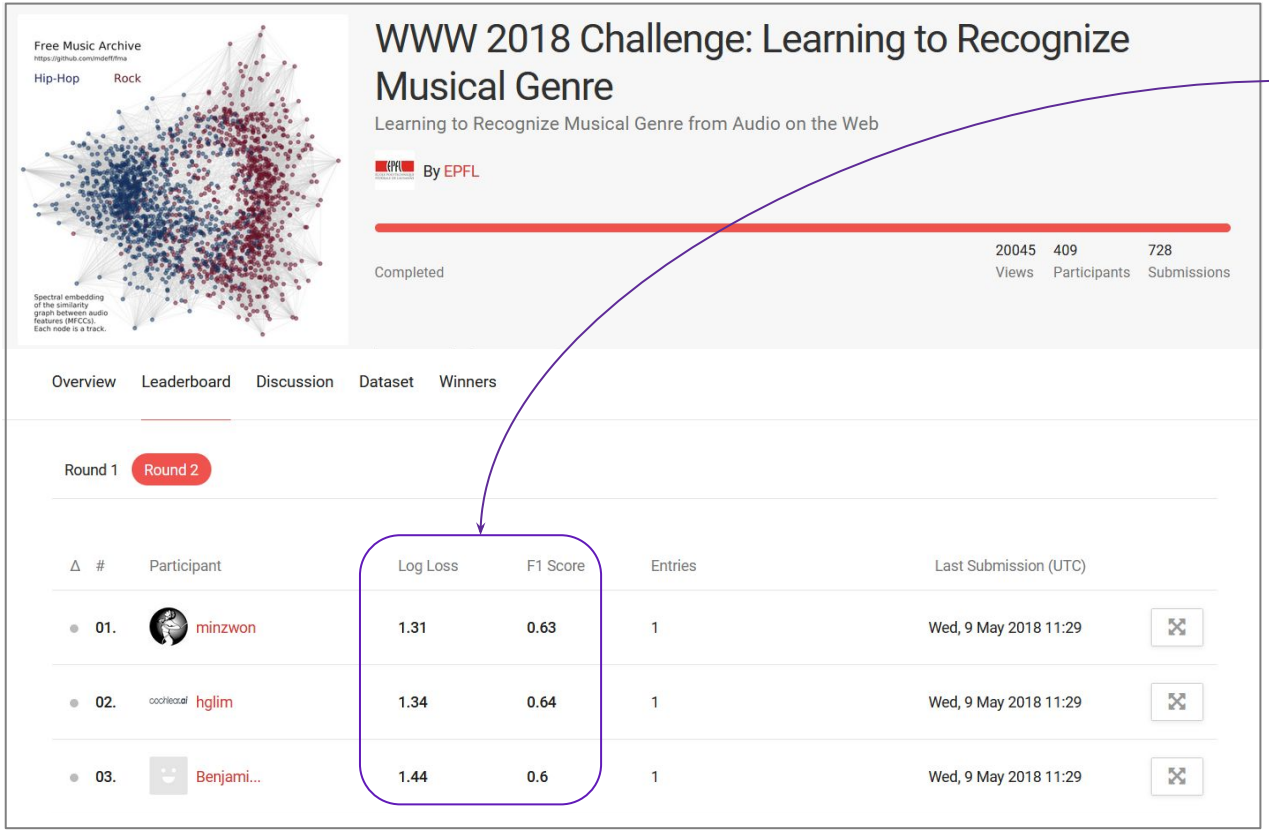
60% accuracy reference line

Transfer Learning Notes:

For these models, inputs were duplicated to match the 3 channel input required by the CNNs pre-trained on ImageNet.

Additional sets of subsampled data was fed into trained models for further fine-tuning, but no improvement in results.

Comparison of Results



Low Accuracy on the FMA

Low accuracy in genre predictions for the dataset in 2018 challenge.

Possible Reasons

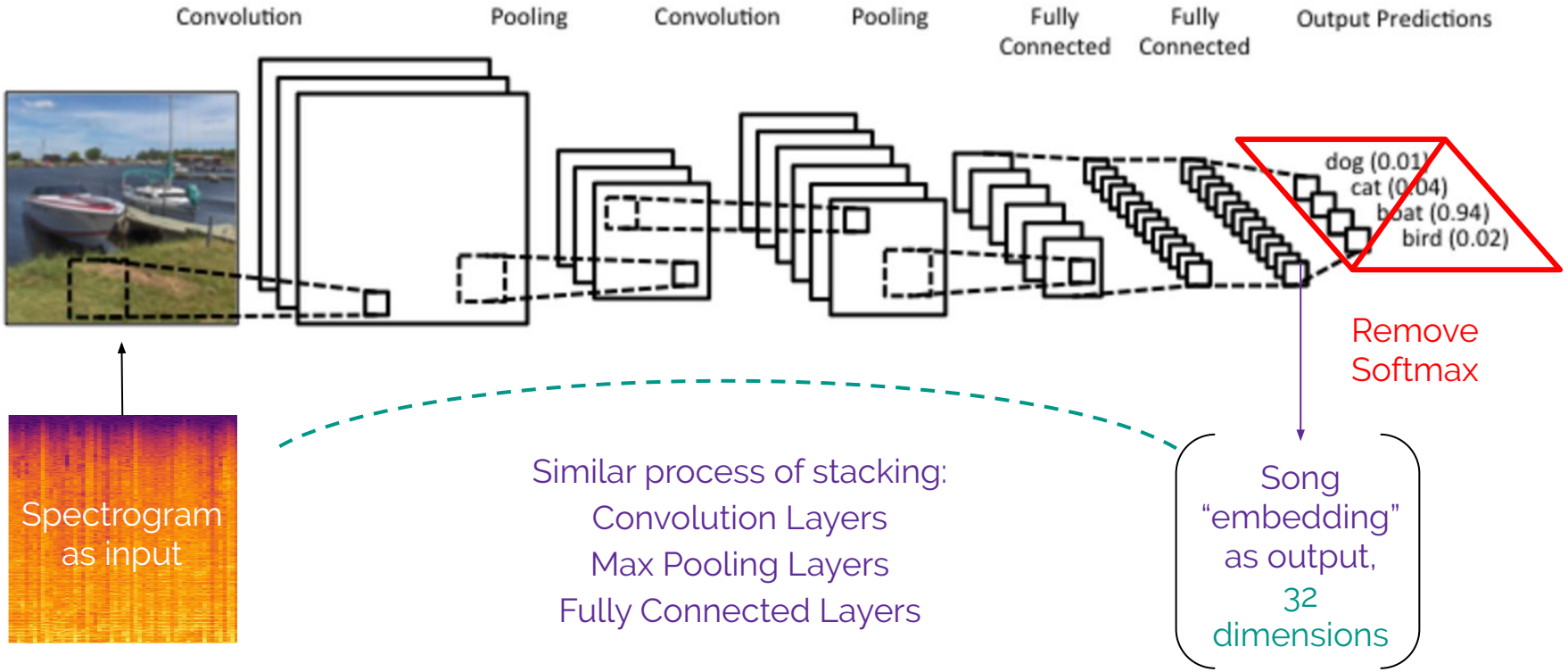
Nature of the types of music and artists on the FMA.

More avant garde, genre-blending songs.

Decided to continue with song embeddings with existing results

Source: https://www.crowdai.org/challenges/www-2018-challenge-learning-to-recognize-musical-genre/leaderboards?challenge_round_id=29

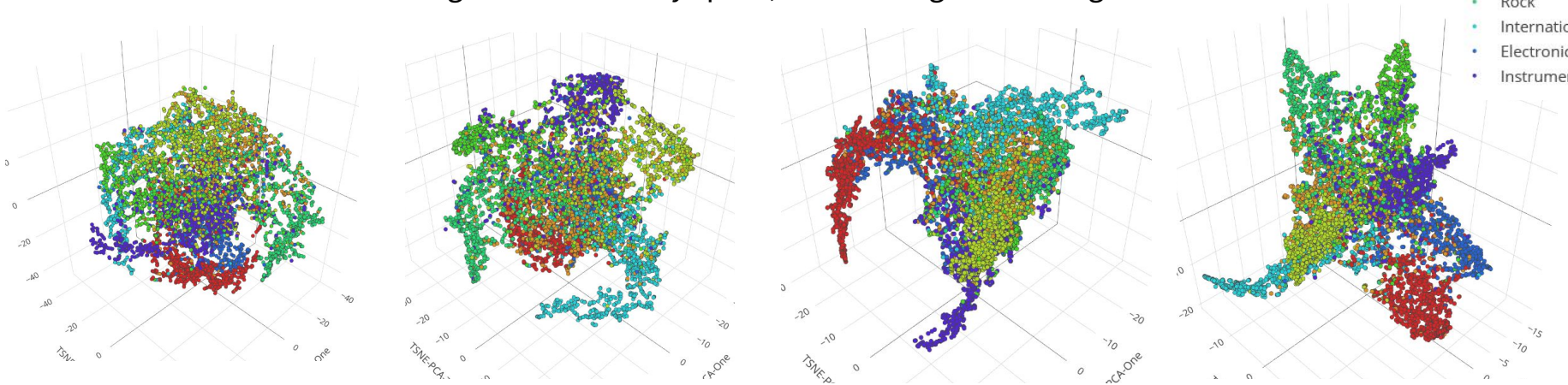
CNN: Song Embeddings



Source: http://www.wildml.com/category/neural-networks/convolutional-neural-networks/

T-SNE 3D Visualization of Embeddings

Function: visualization of high dimensionality space, 32 dim song embeddings



Perplexity: 5 10 15 40

3D scatter plot of songs:

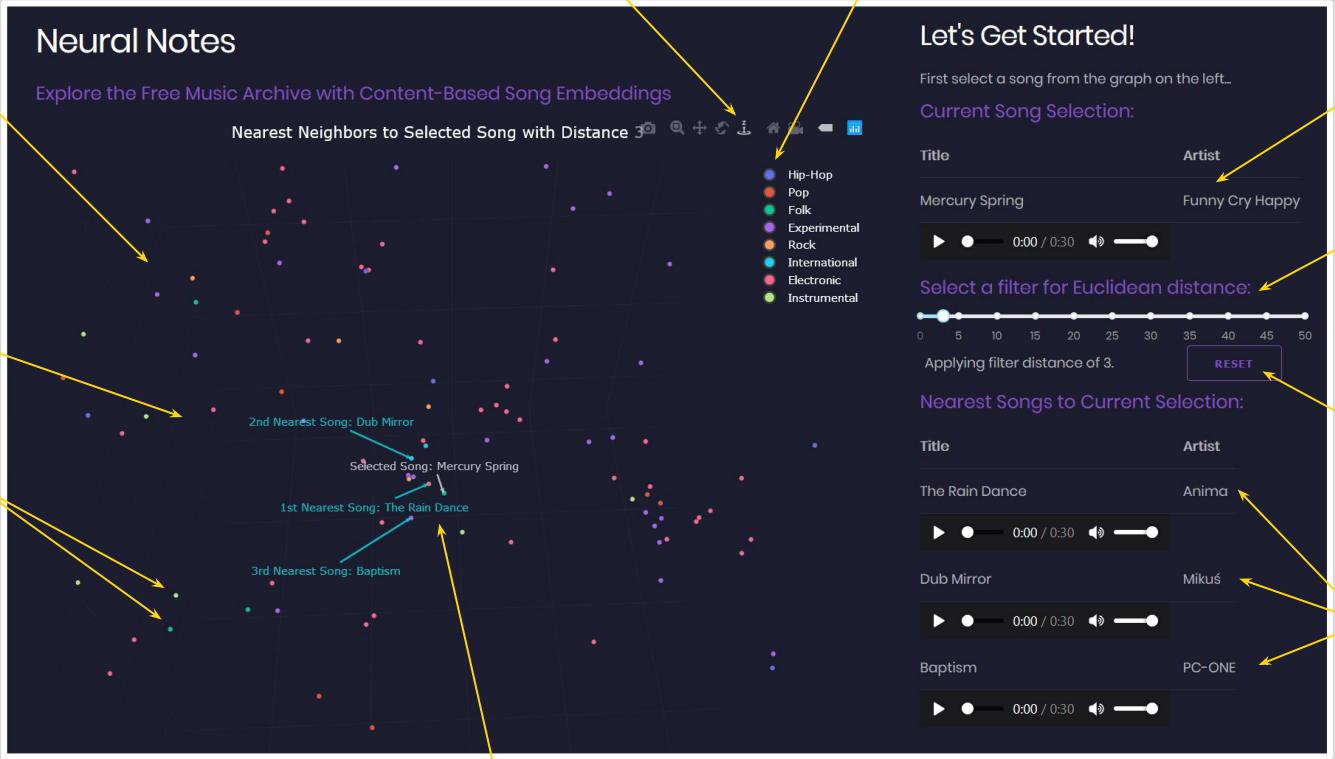
- ➔ To allow for interactivity, and
- ➔ exploration of similar songs through the 3D spatial distance,
- ➔ with spectral song embeddings learned with a convolutional neural network.

Neural Notes: MVP Recap

Control bar for pan, zoom, rotation, etc

Colors indicate different genres, which can be filtered

Interactive 3D space for navigation



Play the current selected song

Slide the distance filter to zoom in on similar songs

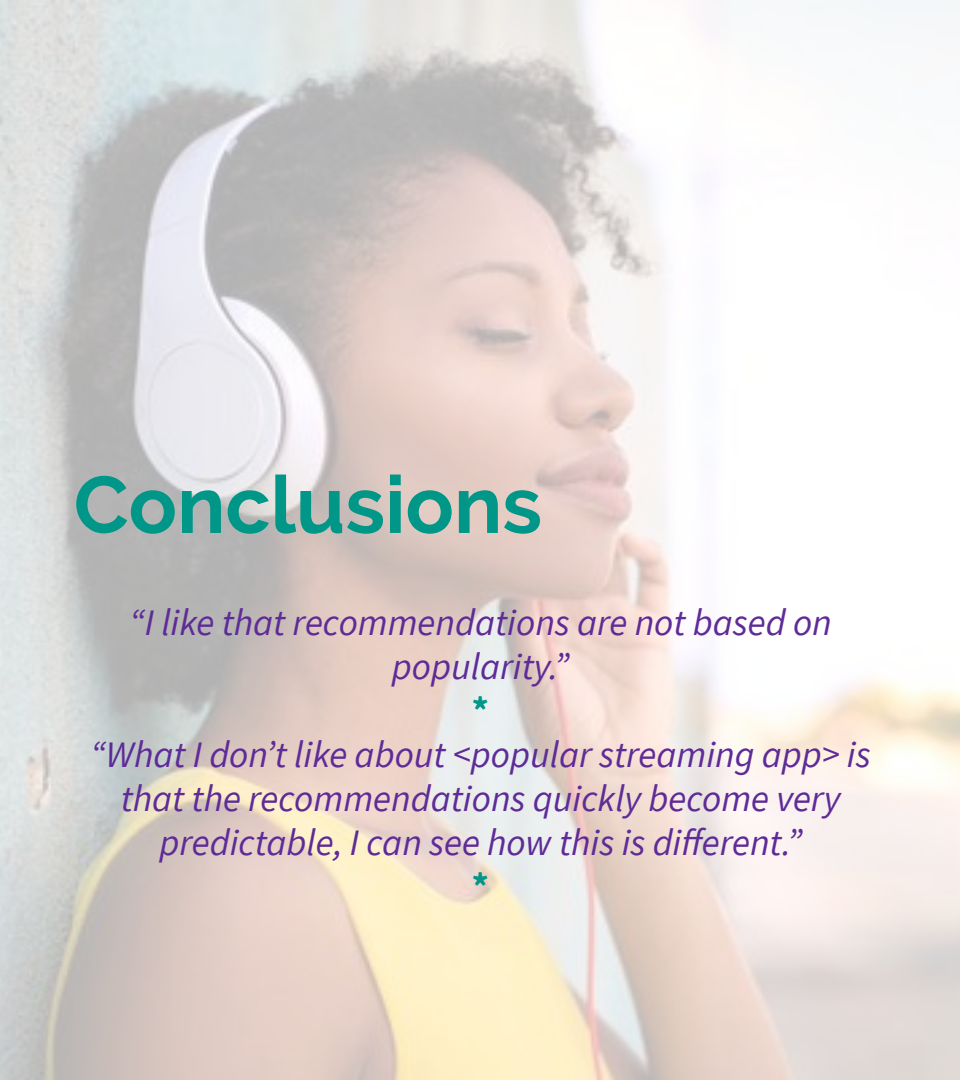
Each song is a point

Reset button for the distance filter

Songs closer to each other are similar based on the learned embeddings

Closest 3 songs to the selected song

Labels for selected song and closest 3 songs



Conclusions

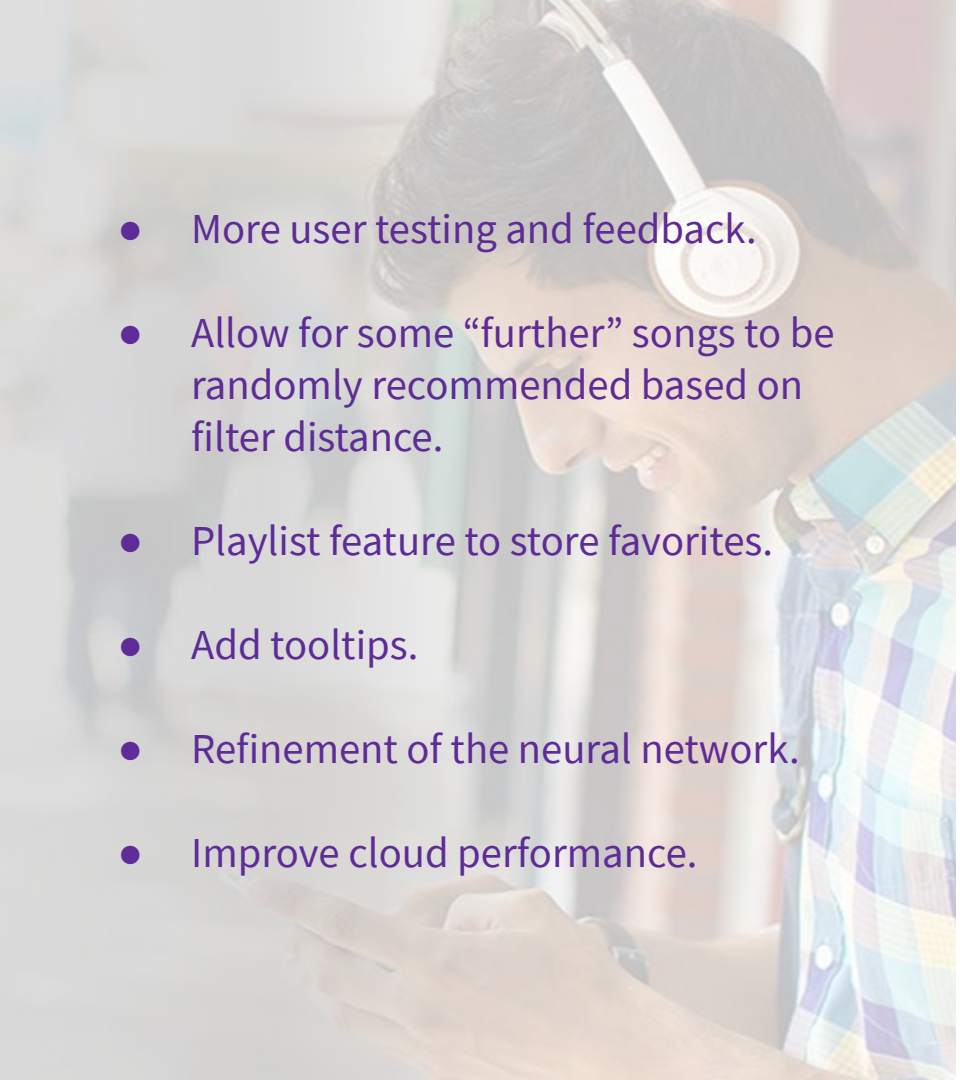
"I like that recommendations are not based on popularity."

*

"What I don't like about <popular streaming app> is that the recommendations quickly become very predictable, I can see how this is different."

*

- Machine learning, deep learning, transfer learning, CNN
- Content-based song profiling using embeddings
- Interactive MVP application for self-directed music "discovery"
- Recommendations without factoring popularity, avoid cold-start problems
- Fun and interesting



- More user testing and feedback.
- Allow for some “further” songs to be randomly recommended based on filter distance.
- Playlist feature to store favorites.
- Add tooltips.
- Refinement of the neural network.
- Improve cloud performance.

Future Work

user feedback
+
additional features
+
better performance

Group Details

mbulkow@berkeley.edu

Madeleine Bulkow-Macy just moved to the Atlanta area and is looking for a job that lets their passion for machine learning solve real-world problems. They joined MIDS in Spring 2018, and are excited to continue learning about machine learning in general and audio processing in particular.

kuangwei@berkeley.edu

Kuangwei Huang runs an engineering department as his day job, and is passionate about developing products that use data science and machine learning to overturn conventions and deliver real world impact. Kuangwei has been a MIDS student since Jan 2018, and while his is excited about the future of data science, he still cringes when asked to write a bio, and hopes an algorithm can do it for him someday.

weixing@berkeley.edu

Weixing Sun is a R&D Scientist in Materials Science. He joined MIDS in Jan 2017 to learn data science techniques and he dreams to use them to increase R&D productivity in the traditional manufacturing and chemical engineering world.

Thank you for listening.

Madeleine Bulkow | Kuangwei Huang | Weixing Sun

References

1. McFee, et Al, 2015, “librosa: Audio and Music Signal Analysis in Python”, 14th Python in Science Conference, Proceedings.
2. Wang A., 2003, “An Industrial-Strength Audio Search Algorithm”, ISMIR 2003, 4th International Conference on Music Information Retrieval, Baltimore, Maryland, USA, October 27-30, 2003, Proceedings.
3. Hershey, et Al, 2016, “CNN Architectures for Large-Scale Audio Classification”, ICASSP 2017.
4. Ebrahimi, Vahabi, 2018, “Predicting Audio Advertisement Quality”, Eleventh ACM International Conference on Web Search and Data Mining, Proceedings.
5. Tan M., et Al, 2019, “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks”, Published in ICML 2019
6. Dieleman S., 2014, “Recommending music on Spotify with deep learning”, http://benanne.github.io/2014/08/05/spotify-cnns.html
7. Guadamuz A., 2015 “Whatever happened to the Long-Tail?”, https://www.technollama.co.uk/whatever-happened-to-the-long-tail
8. Pandey, 2018, “Music Genre Classification with Python”, https://towardsdatascience.com/music-genre-classification-with-python-c714d032f0d8
9. Dwivedi P., 2018, Using CNNs and RNNs for Music Genre Recognition, https://towardsdatascience.com/using-cnns-and-rnns-for-music-genre-recognition-2435fb2ed6af
10. EPFL WWW 2018 Challenge: Learning to Recognize Musical Genre, https://www.crowdai.org/challenges/www-2018-challenge-learning-to-recognize-musical-genre
11. U.S. recorded music revenues RIAA, https://www.riaa.com/u-s-sales-database/
12. Global Music report IFPI 2019, https://www.ifpi.org/news/IFPI-GLOBAL-MUSIC-REPORT-2019

Links

Neural Notes Web App
https://neural-notes.appspot.com/

Informational Webpage
http://people.ischool.berkeley.edu/~weixing/w210_website_v2.1/

UC Berkeley ISchool Project Page
https://www.ischool.berkeley.edu/projects/2019/neural-notes

Github Repo
https://github.com/kuangweihuang/MIDS_w210_capstone_mb_kh_ws