NYU

Relating **Chords** to **Emotions** with Neural Networks

DL4M Final Project

Jingyue Huang (jh8522) & Lydia Morneault (lmm8719)

2023.5.3

PART 01

Project Introduction

Chords & Emotion

Harmony

Harmony, together with rhythm and melody, was thought as able to **elicit emotions** since ancient times.

Consonant harmonies are usually associated with happiness, tranquility, serenity, while dissonant complex harmonies are related with negative emotional states, e.g., tension and sadness, due to the instability they create in the piece.

Major & Minor modes

Major modes have been frequently related with **positive emotions** (e.g., happiness), while **minor modes** are linked to **negative ones** (e.g., sadness).

This emotional association is “neither due to the summation of interval effects nor simply arbitrary, learned cultural artifacts, but rather that harmony has a psychophysical basis dependent on three-tone combinations”.

R. Panda, R. M. Malheiro, and R. P. Paiva, “Audio features for music emotion recognition: A survey”

Project Outcome

Step 1: Automatic Chord Recognition

- Use an existing system to predict the time-varying chord progress

Step 2: Music Emotion Recognition

- Perform transfer learning to estimate the emotion values (valence and arousal)

Step 3: Analyze the relationships

- major-minor tonality V.S. valence-arousal values
- chord recognition model errors V.S. emotional content

This analysis may further assist the improvement of chord recognition models by incorporating useful emotional information!

Dataset - RWC Pop

- **100 3-minute pop music tracks** from RWC Music Database, with time-varying chord labels.
- Each song is split into **10-second clips** for emotional values prediction (the emotion of a song may change over time)
- Use the chord intervals provided in the dataset to match the chord labels with their corresponding emotional values.

Humphrey, E.J., & Bello, J.P. (2015). Four Timely Insights on Automatic Chord Estimation.

PART 02

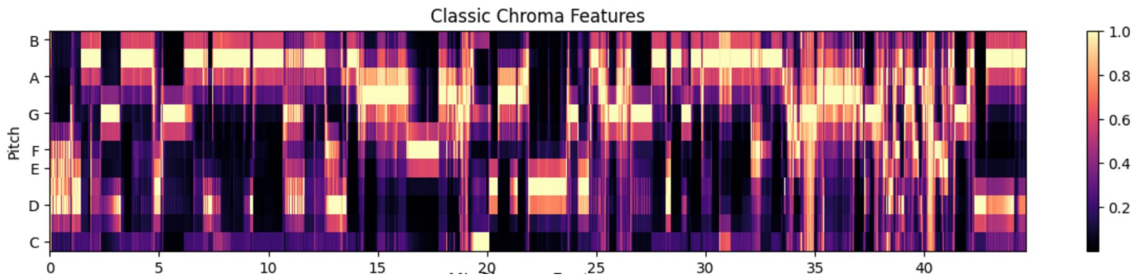
Chord Recognition

madmom

- **madmom** - an audio signal processing library
  - Contains built-in, pre-trained machine learning models
- **DeepChromaProcessor** - computes chroma vectors from audio file using DNN
- **DeepChromaChordRecognitionProcessor** - labels chords with start and end times given audio chroma vectors

Humphrey, E.J., & Bello, J.P. (2015). Four Timely Insights on Automatic Chord Estimation.

Chord Recognition Process



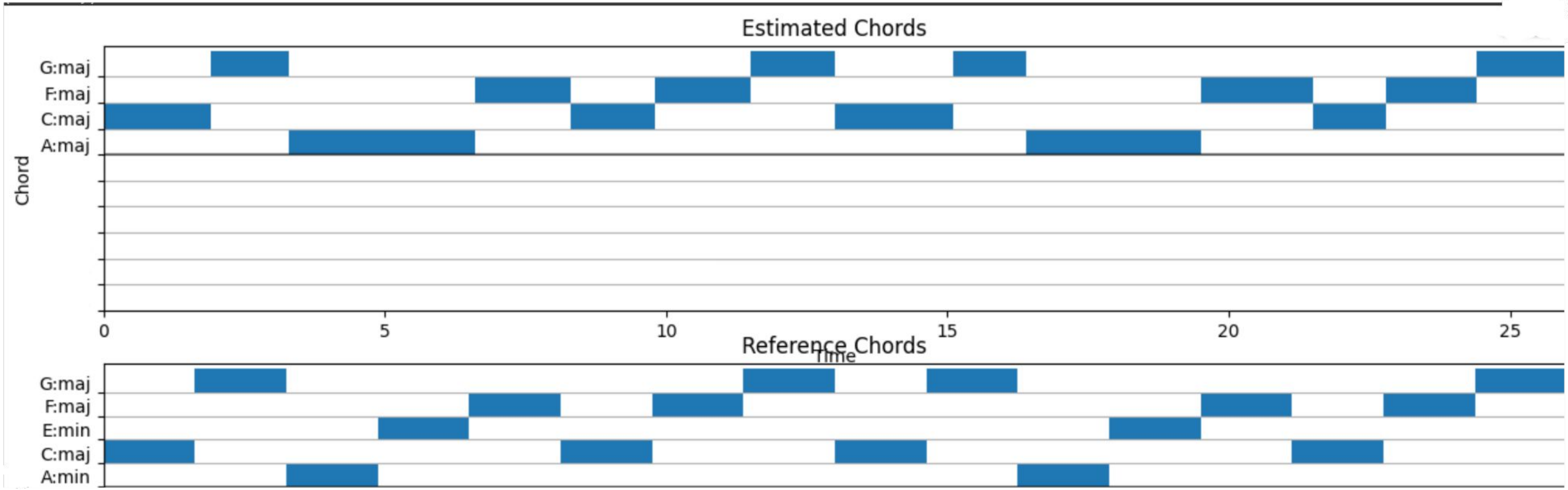
1. Calculate chroma vectors

2. Estimate chords and time intervals from madmom

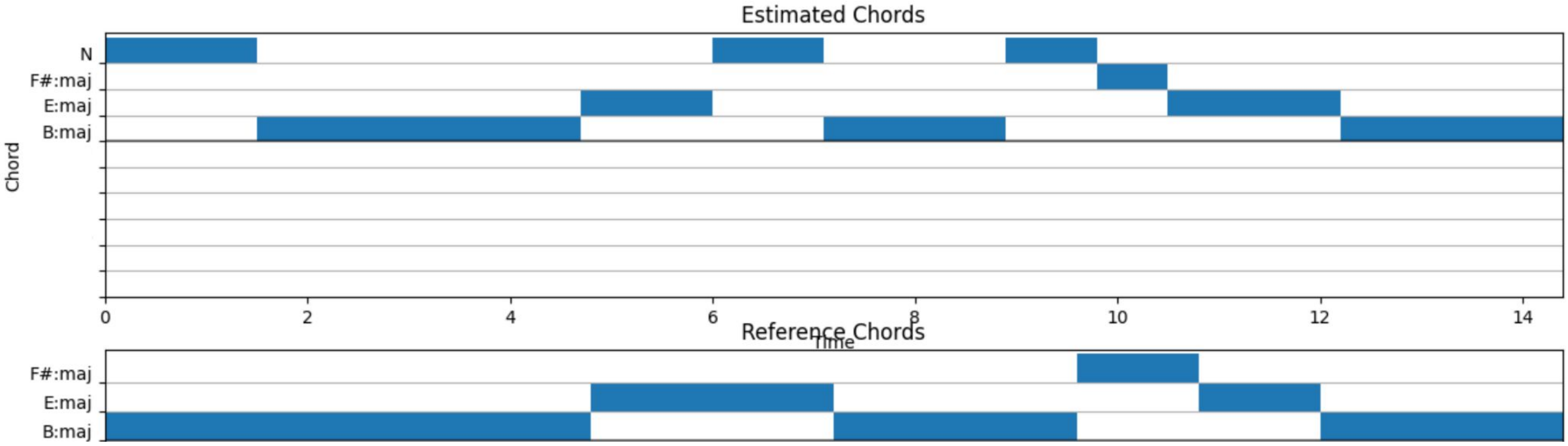
3. Compare accuracy of the true labels to the estimations

NYU

Results - Bad Example



Results - Good Example



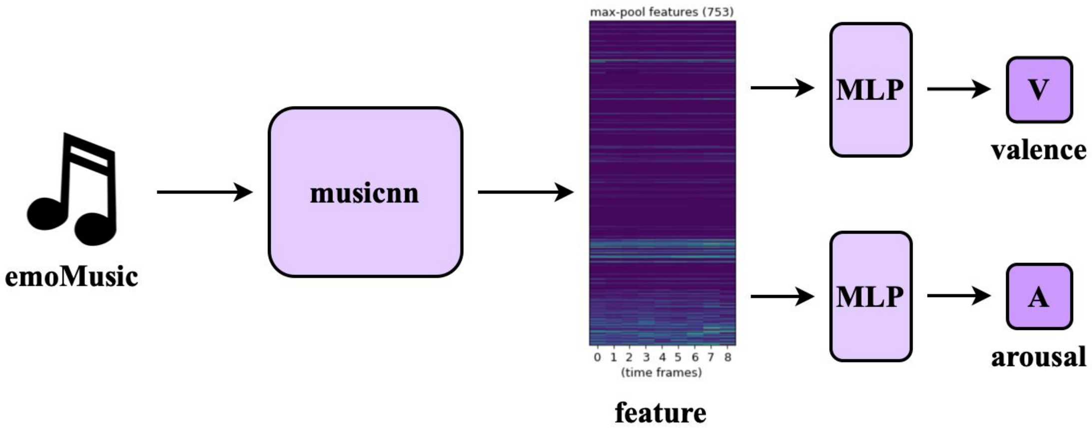
NYU

PART 03

Emotion Recognition

Transfer Learning

- There are **limited data** available for emotion recognition
- Apply pre-trained model designed for MIR tasks for **zero-shot** emotion recognition of RWC Pop



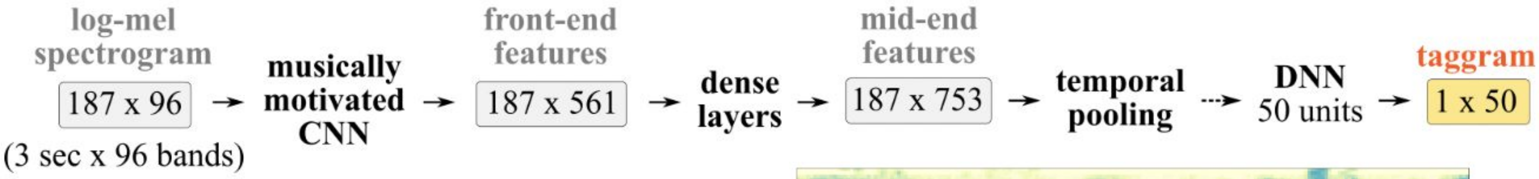
emoMusic

- An **emotion recognition dataset** consisting of 744 songs covering different kinds of mainstream western music.
- Each sample contains 45-second music clips
- Emotion annotation
  - **Valence level**: positive versus negative emotion
  - **Arousal level**: emotional intensity

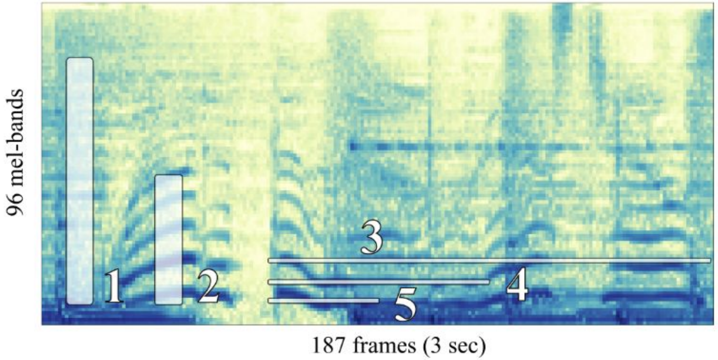
Mohammad Soleymani, Micheal N. Caro, et al. 1000 songs for emotional analysis of music.

musicnn

- A pre-trained musically motivated convolutional neural networks for music audio tagging.



- **Musically motivated CNN**: timbral and temporal features



Jordi Pons, Xavier Serra. musicnn: Pre-trained convolutional neural networks for music audio tagging

Evaluation

- Use **R2 statistics** and root-mean-square error for model evaluation
  - Valence: 0.414
  - Arousal: 0.685
- Our results are close to the values reported in **JukeMIR**, who also used the features extracted from musicnn for emotion recognition on dataset emoMusic as baseline.
  - Valence: 0.466
  - Arousal: 0.703

Rodrigo Castellon et al, Codified Audio Language Modeling Learns Useful Representations For Music Information Retrieval.

Now, Estimate the emotion of RWC

Example 1 - High Valence, High Arousal

- Valence: 6.050
- Arousal: 8.593

Example 2 - Low Valence, Low Arousal

- Valence: 3.900
- Arousal: 2.847

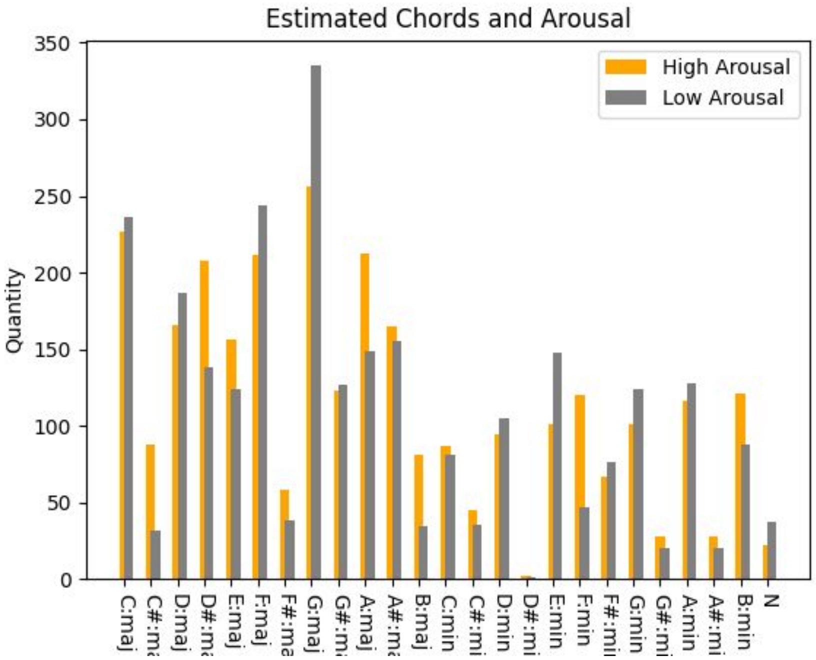
Example 3 - Next Clip of Example 2

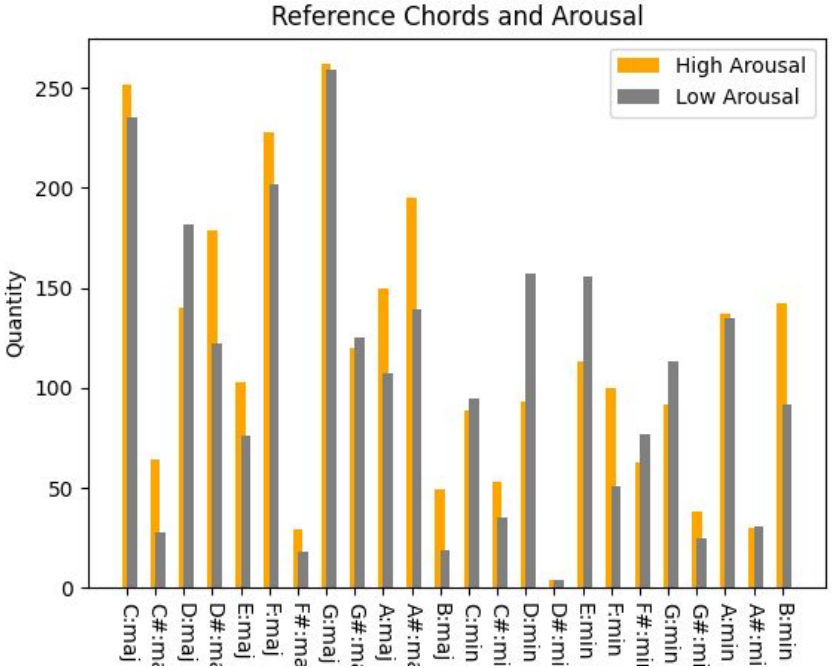
- Valence: 5.777
- Arousal: 4.724

PART 04

Relationship Analysis

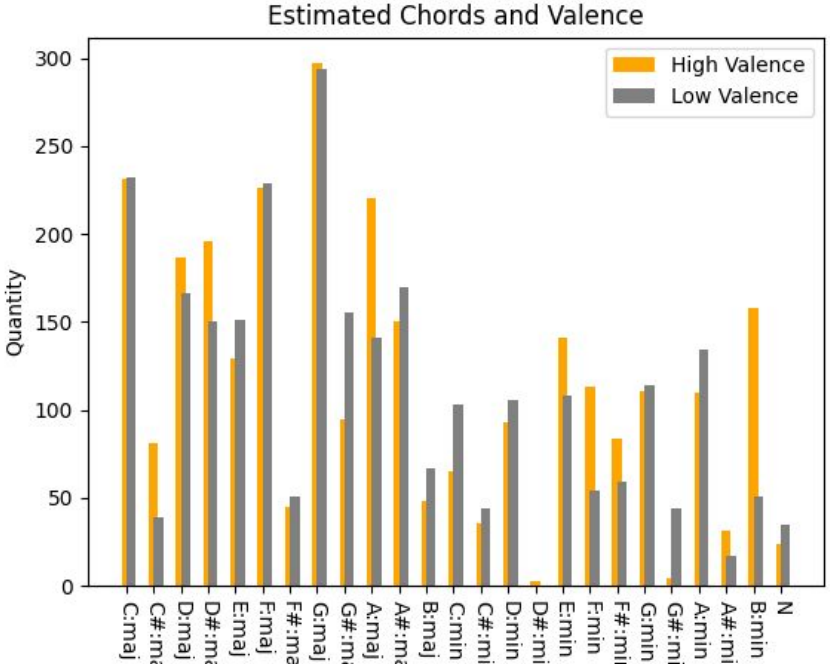
Frequency of High & Low Arousal by Chord

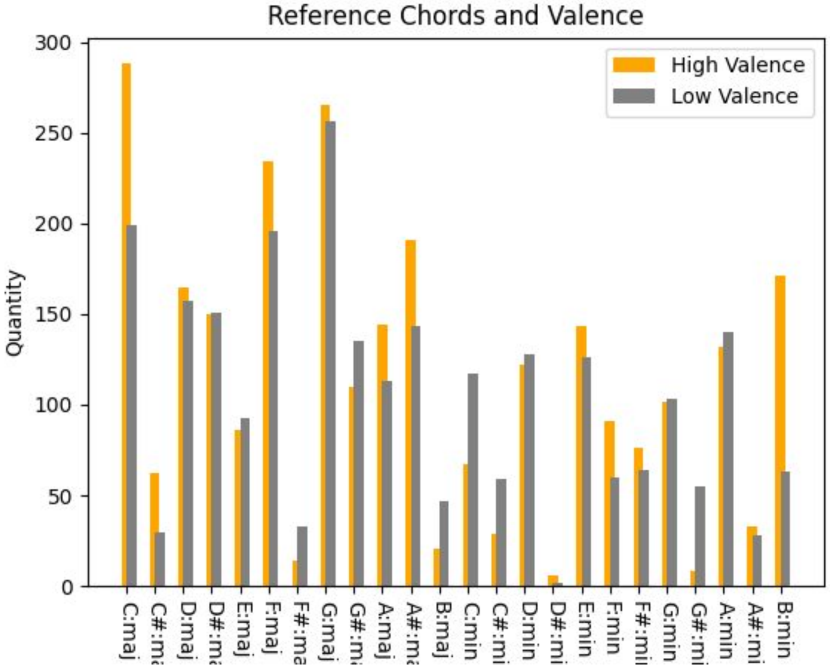




NYU

Frequency of Positive & Negative Valence by Chord



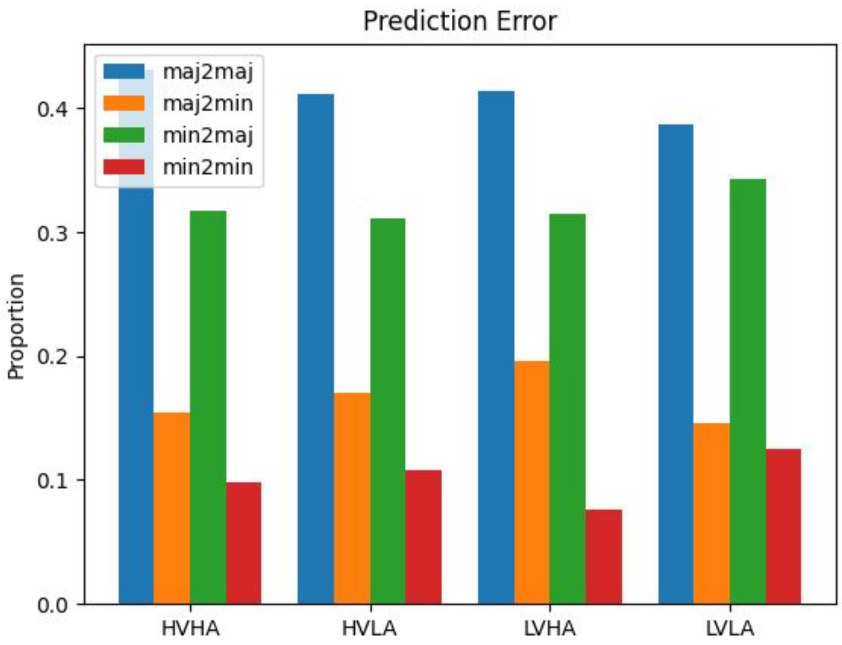


NYU

Findings

- **Reference labels**
  - Major tonality occurs more frequently in high valence/arousal, while minor tonality is associated with low valence/arousal.
  - This pattern is more distinct in the arousal plot as compared to the valence plot.
- **Estimated labels**
  - Chords have a similar frequency across different emotions
  - Many major chords appear more frequently in emotions with low valence or low arousal, which contradicts the findings in psychology.
- **Incorporating emotion information may be useful in chord recognition by assigning varying weights to the prediction of major/minor tonality.**

Analysis of Prediction Errors



- The main source of error is the **misclassification of major chords**.
- The LVHA class has the highest proportion of maj2min errors while the LVLA class has the lowest proportion, suggesting that we could **incorporate high arousal information** to reduce the frequency of misclassified minor chords.
- Valence information may be less useful in improving chord recognition accuracy.

Future Works

- **Better models**
  - Better chord recognition systems considering chord structures have been put forward to improve the classification accuracy.
- **More diverse datasets**
  - Different music styles may appear different chord progress patterns.
- **Larger vocabularies**
  - Chord detection in a small vocabulary (major/minor vocabulary) has been well studied recently, so more efforts should be made to larger vocabularies, such as triads, sixths, sevenths, and suspended chords.
- **More detailed analysis!**

Thank you for listening!

NYU