Machine Learning for Music

Course outline

Machine learning for music

By the end of this course, you will be able to create music compositions just like this!

https://github.com/chrisdonahue/nesmdb/raw/master/static/nlm_generated.mp3

Some main sources for course material...

- **“Intro to Computer Music” (CMU):**

  https://www.cs.cmu.edu/~15322/

- **“Generative AI for Music and Audio Creation” (Michigan):**

  https://hermandong.com/teaching/pat498_598_fall2024/

- **“Deep Learning for Music Analysis and Generation” (NTU):**

  https://github.com/affige/DeepMIR

- **“Computer Music” (UCSD Music):**

  http://musicweb.ucsd.edu/~trsmyth/intro171/intro171.html

- **“Fundamentals of Music Processing (Meinard Müller, International Audio Laboratories Erlangen)**

  https://www.audiolabs-erlangen.de/resources/MIR/FMP/C0/C0.html

- **“Music Classification: Beyond Supervised Learning, Towards Real-world Applications” (Minz Won, Janne Spijkervet, Keunwoo Choi)**

  https://music-classification.github.io/tutorial/landing-page.html

Some other important acknowledgments…

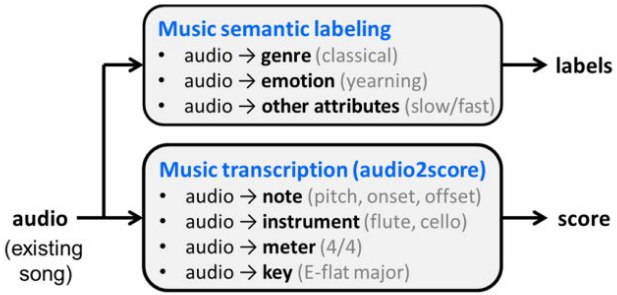
Haven Kim (who wrote many of the HW exercises!)

Zachary Novack (who'll deliver the Module 4 content!)

Amit Namburi, Jingyue Huang, Xin Xu, Haven Kim (various slides)

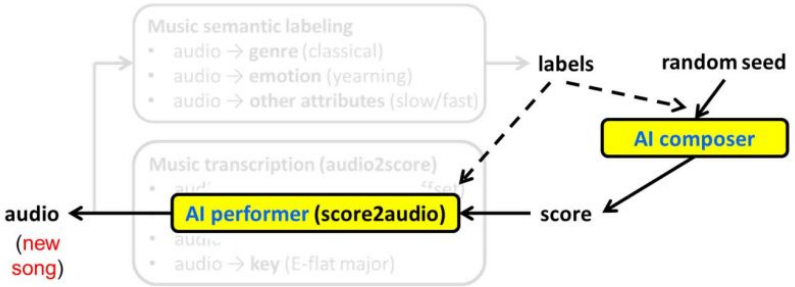
What is Machine Learning for Music?

Music analysis



- Music recommendation (Module 5)
- Music understanding
- Music (Information) Retrieval (Module 2)

Music generation

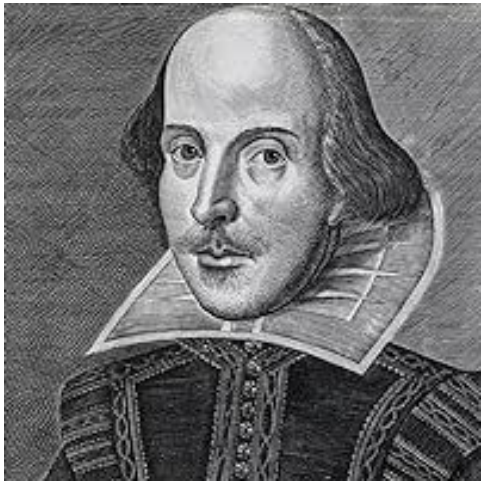


- Symbolic music generation (Module 3)
- Audio generation (Module 4)

Why am *I* interested in ML for music?

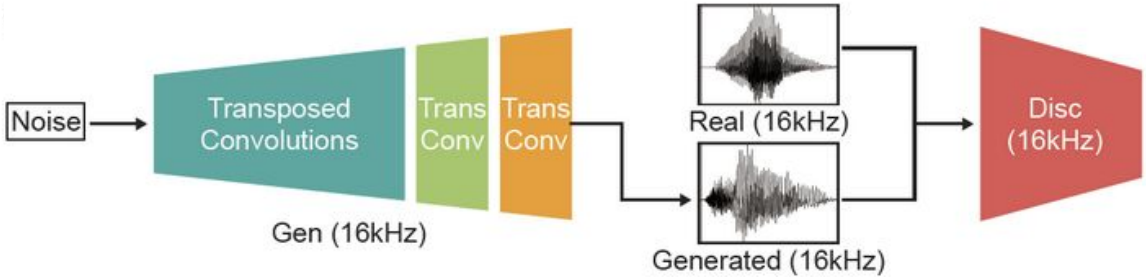
- Have played music most of my life
- Many research directions have recently “converged” (due to e.g. language models, Transformers, diffusion approaches), such that music research overlaps heavily from ideas from language, vision, etc.
- I (and a few others) have been growing a music research group for several years: https://ucsd-musaic.github.io/
- My music students have started to graduate and I was jealous of the classes they were teaching

“The man that hath no music in himself, Nor is not moved with concord of sweet sounds, Is fit for treasons, stratagems, and spoils; The motions of his spirit are dull as night, And his affections dark as Erebus. **Let no such man be trusted.** Mark the music.”



Some of our lab's work on ML for music

WaveGAN (Donahue *et al.*, 2019)

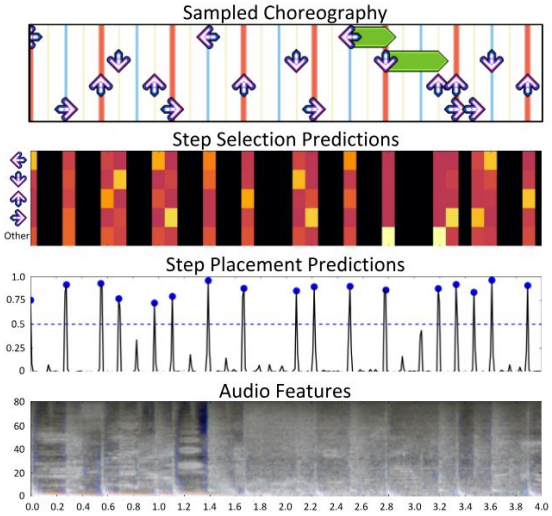


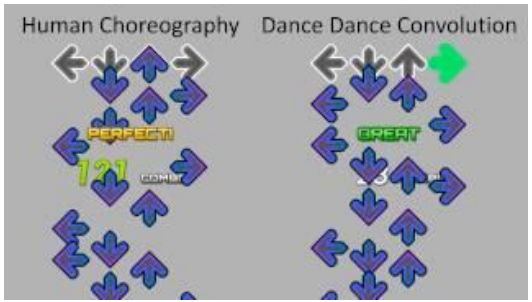
- First attempt to apply GANs for synthesizing audio (e.g., drums, speech)
- What is a GAN?
  - A Generator tries to create "fake" data (audio) that looks real
  - A Discriminator tries to figure out which data (audio) is real and which is fake
  - As they compete, both get better at their jobs: the Generator gets better at creating convincing fakes, and the Discriminator gets better at spotting them
- Sound examples: https://chrisdonahue.com/wavegan_examples/

Some of our lab's work on ML for music

Dance Dance convolution (Donahue *et al.*, 2017)

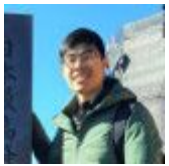
- Proposed and presented the task of ***learning to choreograph***.
- What is *learning to choreograph*?: Learning to produce a step chart (a sequence of arrow steps that players need to hit during a song) in Dance Dance Revolution given music audio
- Video: https://www.youtube.com/watch?v=yUc3O237p9M





Some of our lab's work on ML for music

Multitrack music transformer (Dong *et al.*, 2023)



- Generating symbolic music with multiple instruments simultaneously (suitable for real-time improvisation)
- What is **symbolic music**? a digital representation of music that uses discrete symbols (like notes, durations, instruments) rather than raw audio waveforms (**Module 3!**)

Some of our lab's work on ML for music

DITTO - Diffusion inference-time T-optimization for music generation (Novack *et al.*, 2024)

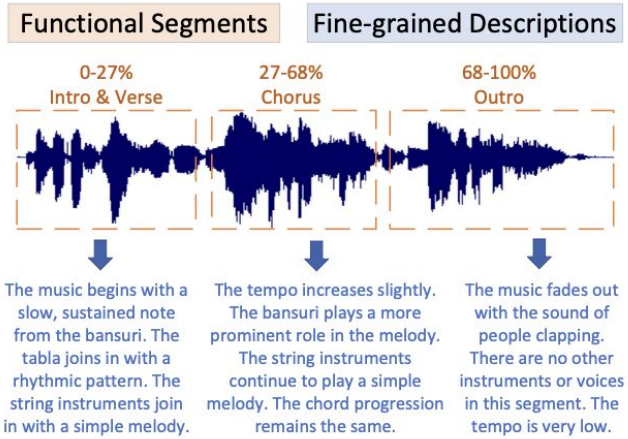
Diffusion model for music generation (**Module 4!**)

Some of our lab's work on ML for music

FUTGA-MIR (Wu *et al.*, 2024)

Enhancing Fine-grained and Temporally-aware Music Understanding with Music Information Retrieval

This is a model that can generate fine-grained, time-aware captions for music, creating new possibilities for music understanding and interaction



Some of our lab’s work on ML for music

Lots more on https://ucsd-musaic.github.io/ !

Why are *you* interested in ML for music?

Why are *you* interested in ML for music?

...

Course outline

The course is broken into 5 “modules”, **roughly:**

- **Module 1:** data structures for music and data ingestion
- **Module 2:** predictive pipelines for music & music information retrieval (MIR)
- **Module 3:** symbolic-domain music generation
- **Module 4:** continuous-domain music generation
- **Module 5:** other topics in music information retrieval

Assessment (see next slide!)

- Homework: **40%**
  - Each module (roughly) will be associated with one homework assignment, worth 10% of your final grade. Your lowest homework grade will be dropped (or you can skip one); most people will probably skip the last. Homework assignments are due in **weeks 2, 4, 6, 8, and 10**
- Midterm: **20%**
  - **Week 6;** covering content from Modules 1-3
- First assignment: **20%**
  - **Week 7;** a “kaggle-like” competition on a MIR task (e.g. chord labeling or genre classification)
- Second assignment: **20%**
  - **Week 10;** a project showcase, featuring a “performance” of, and brief report of generated music by your group (you don’t have to perform the piece yourself!)
  - Groups can be 1-4 students

Assessment

I don’t plan to run the midterm this quarter since the course is running for the first time: I’m not sure how quickly I’ll get through the material or what adjustments I need to make as people succeed or fail at the homeworks; so this year the breakdown will be more like:

- HW: 50%
- Assignment 1: 25%
- Assignment 2: 25%

Assessment

- **Homework** will be **autograded:** you will submit your code and generated outputs and receive grades immediately; you can submit an unlimited number of times
- The **midterm** will be a **take-home**, 12 hour format, mostly to accommodate remote students; the submission process will be the same as for the homework
- The **first assignment** will be mostly graded based on your ability to pass performance thresholds for the given task; a *tiny fraction* of your grade will be based on performance relevant to your peers
- The **second assignment** will be peer graded

Assessment

- **Homework** is meant to be fairly “easy” and just to *make sure everyone is staying on top of the material*
- The **midterm** is intended to make sure everyone has properly *synthesized the course materials* (up to that point); that being said it’s fairly similar in format to homeworks, subject to a time constraint
- The **first assignment** is intended to test your ability to *just get something practical working* (regardless of whether your solution is elegant or not!)
- The **second assignment** is intended to test your ability to *apply what you’ve learned creatively*

Expected knowledge

- **Basic** data processing
  - Text manipulation: count instances of a word in a string, remove punctuation, etc.
  - Process formatted data, e.g. JSON, html, CSV files etc.
  - Install and run libraries to process structured data formats
- **Basic** mathematics
  - Some linear algebra
  - Some optimization
  - Some statistics

The expectation is not that everyone comes in knowing all these things, but that if you don't know them already, you will self-study them; links/resources will be made available

Expected knowledge

No prior experience in machine learning is required

But! Many people in the class (probably?) have some ML background, so I don't want the revision to go too slow for them

Compared to my other classes, this class teaches specific methods much less: we need to rely on fancier models ("deep learning" etc.) which we'll mostly interface with via library implementations

Expected knowledge

No prior experience in music is required!

I'll mostly assume people *don't* know "music stuff," and that you're not planning to learn it this quarter

I'll teach the bare minimum (what is a *scale*, what is a *chord*, etc.) that is necessary to deliver the material

Arguably part of the point of this class is that *musical expression should be accessible to everyone* without requiring deep musical knowledge (through the magic of ML!)

Happy to talk about "theory stuff" to the extent anyone is interested (and to the extent that I know enough myself!)

Course outline in detail

Module 1: Data structures for music and data ingestion

- **Primer on musical and machine learning terminology**
  - Not too much, mostly here as a reference, to make sure any terms are clear as they come up later in the course
- **Symbolic representations of music (e.g. midi/piano roll)**
  - How can music be represented as discrete sequences of events, e.g. as “notes” with associated metadata (timing, velocity, etc.)
  - What are the advantages/disadvantages of different representations?
  - How can we read and ingest datasets?
- **Continuous representations of music (waveforms and spectrograms)**
  - How can *recordings of* music and audio by represented?
  - How can we convert between these formats and those above?
  - What are the advantages and disadvantages of different formats?
- **Case study:** Piano Genie

Course outline in detail

Module 2: Predictive pipelines for music & Music Information Retrieval (MIR)

- Covers classification and "standard" Music Information Retrieval (MIR) tasks, e.g. tagging, genre classification, onset detection, chord labeling, etc.
- Generally speaking: how to implement “standard” ML approaches where the input space is music
- This module also serves as revision of ML fundamentals (linear modeling, markov chains, and very basic introduction to deep learning)
- **Case Study:** Fine-grained music understanding

Course outline in detail

Module 3: Symbolic-domain music generation

- How should music be represented symbolically (i.e., as “tokens”)?
  - Chords, notes, “piano rolls”, structured representations, etc.
- Markov-chain approaches
- Language-modeling approaches to symbolic generation
  - Music as text; music as “tokens”
  - Language-to-music generation
- **Case studies:** MuseCoco; Reviving 15th-Century Korean Court Music

Course outline in detail

Module 4: Continuous-domain music generation

- How should music be represented *continuously* (i.e. as “audio”)?
  - Raw waveforms, spectrograms, audio codecs, etc.
- Old school approaches to audio synthesis
  - Classical DSP / analog synthesis, RNNs, VAEs, GANs
- Modern audio-domain music generation
  - Autoregressive audio generation
  - Diffusion Based Generation
- **Case studies:** MusicGen, Stable Audio Open

Course outline in detail

Module 5: other topics in music information retrieval

- **Music Recommendation**
  - What is a music recommender system and how can we approach them?
  - Dataset for a music recommender system
  - Deep content-based music recommendation (Van den Oord et al., 2013)
- **Playlist Continuation**
  - What is automatic playlist continuation and how can we approach them?
  - RecSys2018 Playlist Continuation Challenge
- **Lyric Translation**
  - How is automatic lyric translation different from standard machine translation?
  - How can we evaluate automatic lyric translation?
  - How can we approach automatic lyric translation?
- **Ethical Considerations of Music AI**
  - Ownership and authorship
  - Inputs, Outputs

Course outline in detail

Guest lectures!!!

- **Prof. Taylor Berg-Kirkpatrick (aka "Synthlord"):** will discuss modular synthesis and give a musical performance (unless he chickens out)
- **Hugo Flores Garcia** (Northwestern University): May 22 (**Thu week 8**) (see e.g. https://www.youtube.com/watch?v=XN4zub7nlhY)
- **Yusong Wu** (MILA/Université de Montréal): May 27 (**Tue week 9**) (see e.g. https://storage.googleapis.com/realchords/index.html)
- exact dates TBD – above is **tentative**
- **Please show up!**

Course outline in detail

Week 10:

Music! You will have the opportunity to perform compositions from your assignments live in front of the class!

This will be purely voluntary and will be worth **zero points** (since I don't want any credit to be available only to musicians ∩ extroverts); **it will just be *for fun***

If nobody volunteers, you'll have to listen to me perform

Course outline in detail

Questions?