Pop Music Hook Generation Using Neural Networks

Daniel Shen ‘22
Advisor: Professor Jeffrey Snyder

In a Nutshell

I trained a neural network to **output eight-measure musical hooks in lead sheet form** using a dataset of U.S. contemporary pop music.

I took one output example from my model, made some tweaks, and **turned it into a full-length pop EDM track, “Will You Learn”.**

Motivation and Goals

Lyrics from “Will You Learn”

They tellin’ me you can learn anything

Drive a car, paint a picture, show the next thing I see

On my feed, babe if you can do all that

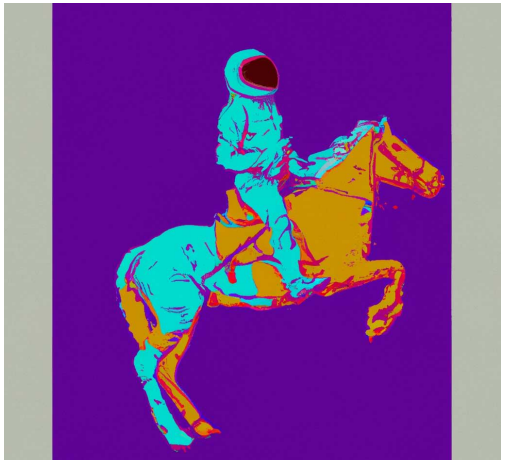
Maybe you can write for me a tune that slaps

Motivation: machine learning is becoming increasingly ubiquitous in our world.

“Drive a car”

Tesla Autopilot

“Paint a picture”



DALL-E 2: “an astronaut riding a horse in the style of Andy Warhol”

“Show the next thing I see on my feed”

Facebook News Feed

Project Goals

Design a deep learning-based program for generating eight-bar pop music hooks.

Investigate the capabilities of deep learning technology and use it to create a compositional tool for songwriters/producers.

Crash Course in Neural Networks

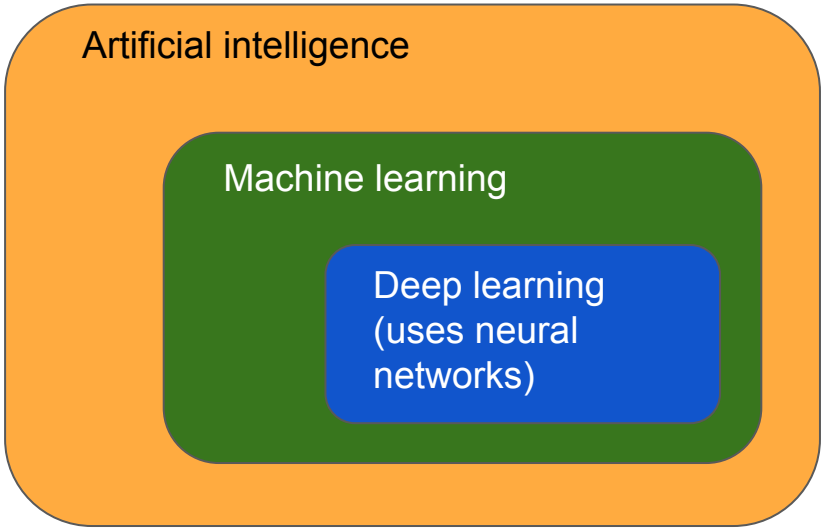
Clarifying Key Terms

Artificial intelligence (AI): any intelligent behavior exhibited by a machine

Machine learning (ML): using lots of data to train AI in machines

Neural network (NN): specific type of computational model used for ML

Deep learning (DL): ML using neural networks (often multiple network layers)



The mathematical foundation of NNs is *linear regression.*

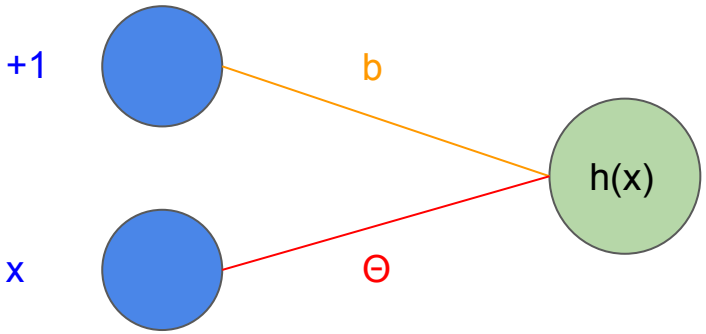
Example: given the **square footage** of a house, determine its **price**


```
price = b + m * (sq. footage)

y     = b + m * x

h(x)  = b + Θx
```

Goal: "tune" the values of b and Θ to best fit the real-world data
(gradient descent etc.)

The mathematical foundation of NNs is *linear regression.*

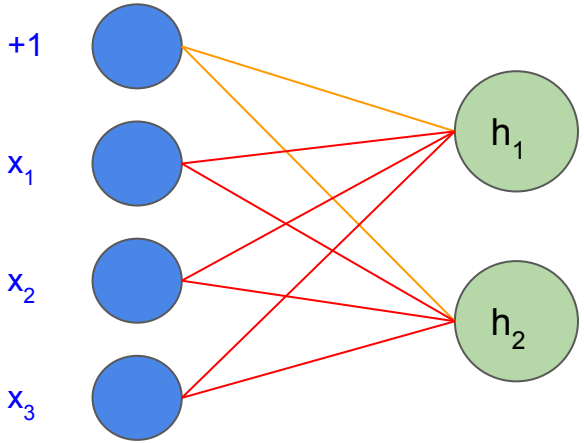
Example: given the **square footage** of a house, determine its **price**



$$h(x) = b + \Theta x$$

The mathematical foundation of NNs is ***multivariate*** *linear regression.*

Example: given the **square footage**, **# bedrooms**, and **age** of a house, determine its **price** AND **average length of homeownership**

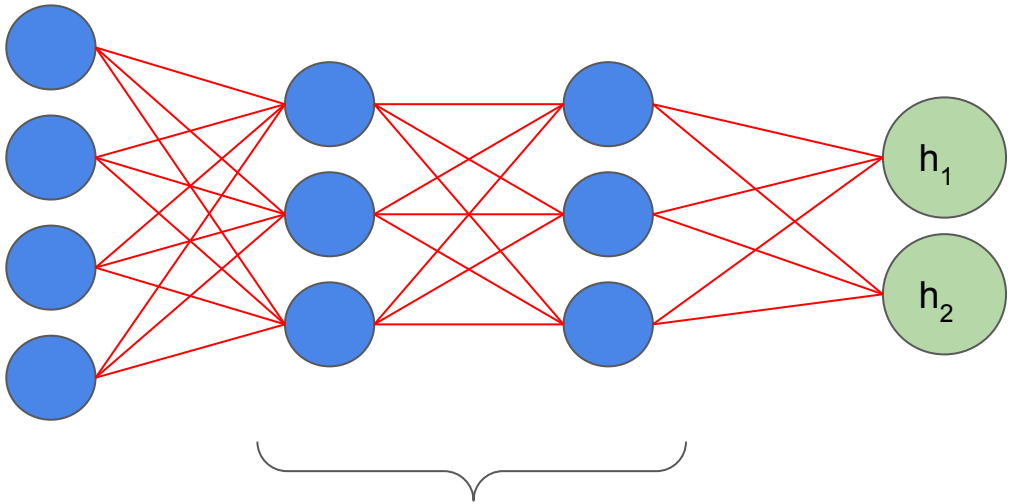


$$h(x) = b + \Theta x$$

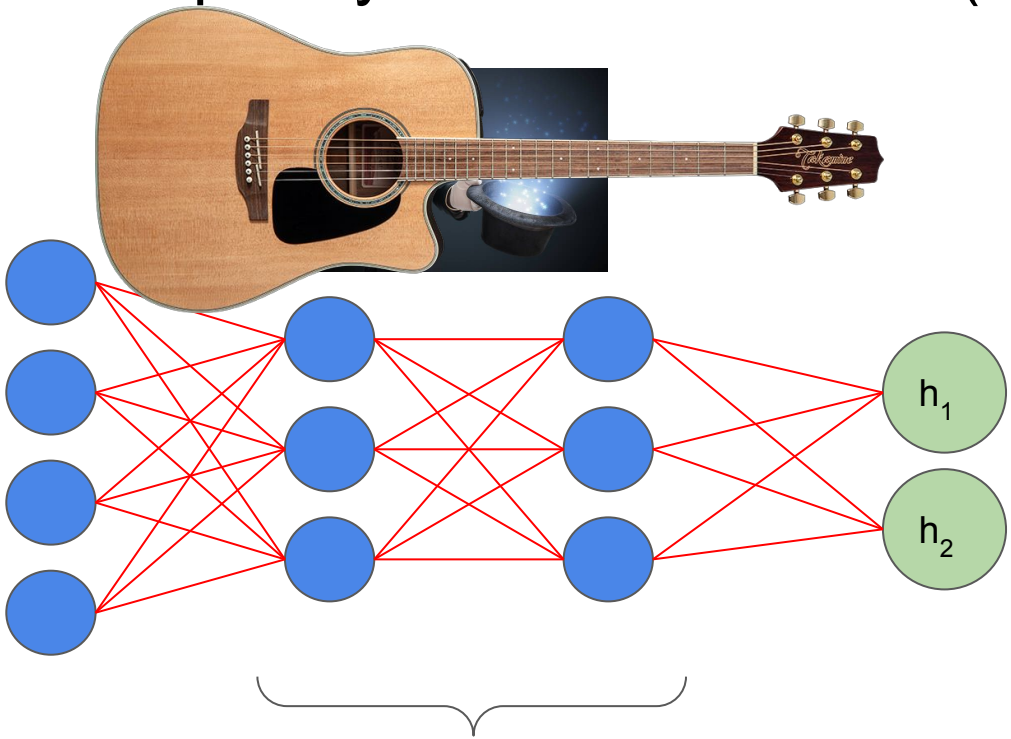
h, b, and x are vectors. Θ is a 2 x 3 matrix

Goal: "tune" the values of b and Θ to best fit the real-world data

You can add multiple layers to the network! (This is deep learning.)

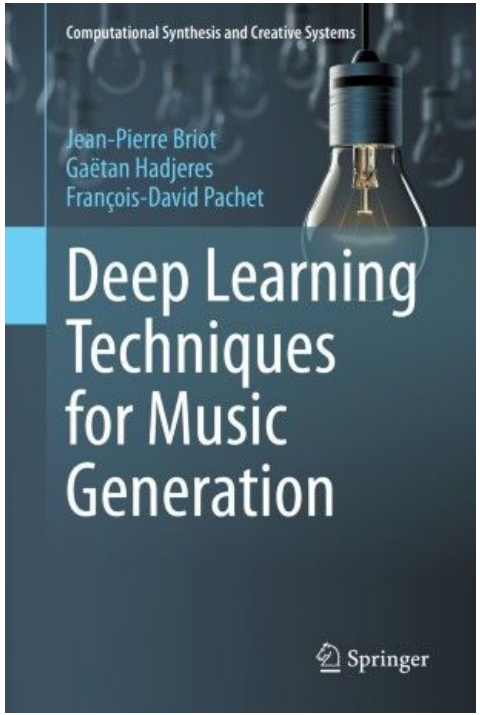


You can add multiple layers to the network! (This is deep learning.)



Problem Context and Related Work

DL-based music generation is still in its infancy.



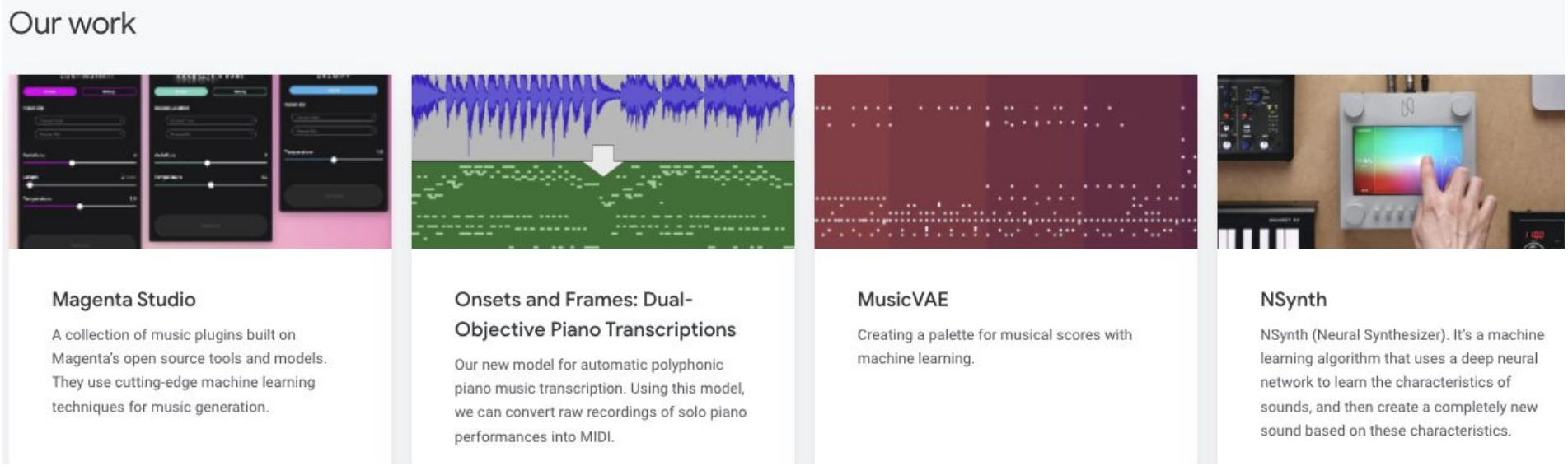
First comprehensive lit review of deep learning-base music generation.

Motivated out of “the lack … of a comprehensive survey and analysis of this active research domain” (p. 6)

Published in 2020! (Earlier draft in 2017)

DL-based music generation is gaining popularity.

Google Magenta: “an open source research project exploring the role of machine learning as a tool in the creative process.”



Current literature spans a wide range of generation objectives.

- Lead sheet generation
  - C. De Boom, S. Van Laere, T. Verbelen, and B. Dhoedt, “Rhythm, chord and melody generation for lead sheets using recurrent neural networks,” Machine Learning and Knowledge Discovery in Databases, pp. 454–461, 2020.
- Generating variations on a theme
  - P. Roy, A. Papadopoulos, and F. Pachet, “Sampling variations of lead sheets,” arXiv preprint arXiv:1703.00760, 2017.
- Generation of chords/harmony given an input melody
  - Y.-C. Yeh, W.-Y. Hsiao, S. Fukayama, T. Kitahara, B. Genchel, H.-M. Liu, H.- W. Dong, Y. Chen, T. Leong, and Y.-H. Yang, “Automatic melody harmonization with triad chords: A comparative study,” Journal of New Music Research, vol. 50, no. 1, pp. 37–51, 2021.
- MIDI performance generation (including note velocities and micro-timings)
  - S. Oore, I. Simon, S. Dieleman, D. Eck, and K. Simonyan, “This time with feeling: Learning expressive musical performance,” Neural Computing and Applications, vol. 32, no. 4, pp. 955–967, 2020.
- Raw audio waveform generation
  - A. v. d. Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu, “Wavenet: A generative model for raw audio,” arXiv preprint arXiv:1609.03499, 2016.

Current literature spans a wide range of generation objectives.

- **Lead sheet generation**
  - **C. De Boom, S. Van Laere, T. Verbelen, and B. Dhoedt, “Rhythm, chord and melody generation for lead sheets using recurrent neural networks,” Machine Learning and Knowledge Discovery in Databases, pp. 454–461, 2020.**
- Generating variations on a theme
  - P. Roy, A. Papadopoulos, and F. Pachet, “Sampling variations of lead sheets,” arXiv preprint arXiv:1703.00760, 2017.
- Generation of chords/harmony given an input melody
  - Y.-C. Yeh, W.-Y. Hsiao, S. Fukayama, T. Kitahara, B. Genchel, H.-M. Liu, H.- W. Dong, Y. Chen, T. Leong, and Y.-H. Yang, “Automatic melody harmonization with triad chords: A comparative study,” Journal of New Music Research, vol. 50, no. 1, pp. 37–51, 2021.
- MIDI performance generation (including note velocities and micro-timings)
  - S. Oore, I. Simon, S. Dieleman, D. Eck, and K. Simonyan, “This time with feeling: Learning expressive musical performance,” Neural Computing and Applications, vol. 32, no. 4, pp. 955–967, 2020.
- Raw audio waveform generation
  - A. v. d. Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu, “Wavenet: A generative model for raw audio,” arXiv preprint arXiv:1609.03499, 2016.

Narrowing the Problem Scope

I only consider U.S. contemporary pop music.

Specifically, I only consider a subset of U.S. pop with certain attributes:

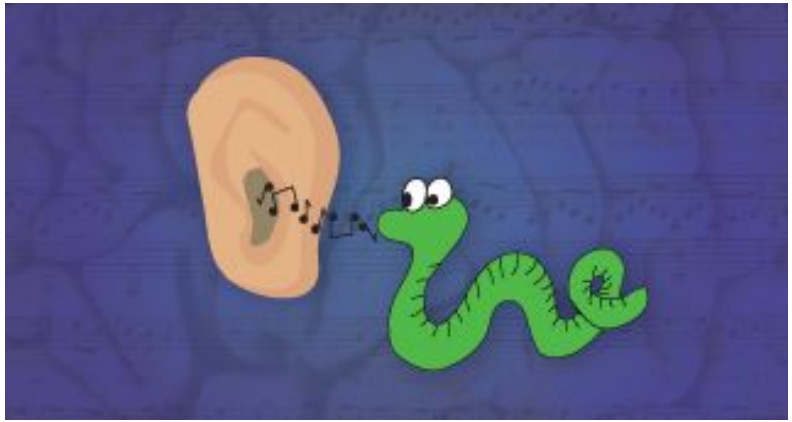
- 4/4 time
- No changes in meter or key signature
- Features prominent melodic component
- Melody range < 2 octaves
- Other attributes (stay tuned!)

I only consider *hooks*, not entire songs.

Working definition: eight-bar self-contained unit of pop music with a memorable tune, usually synonymous with chorus***

Long-term structure is difficult to learn.

Hooks are meant to be catchy, so they are a high-impact area to focus on.

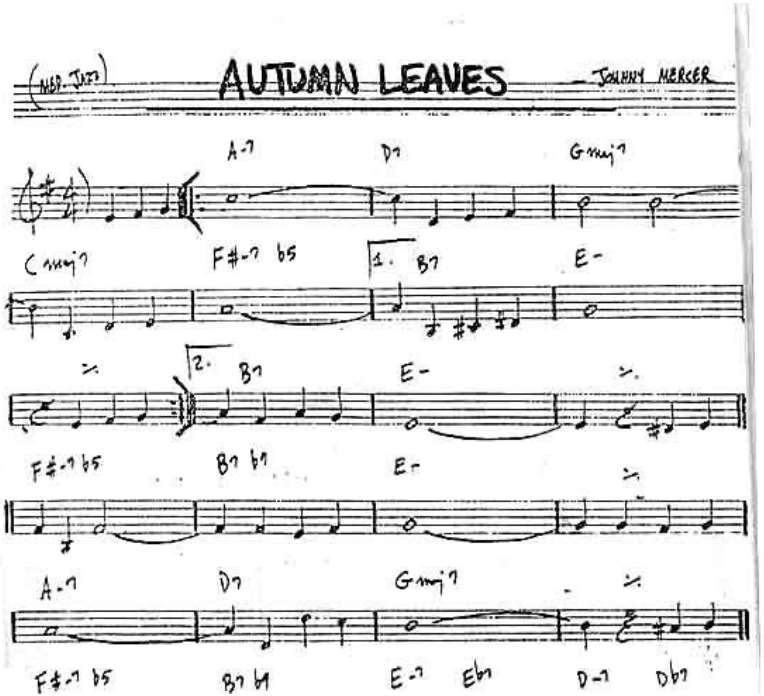


I represent hooks in a symbolic lead-sheet-like form.

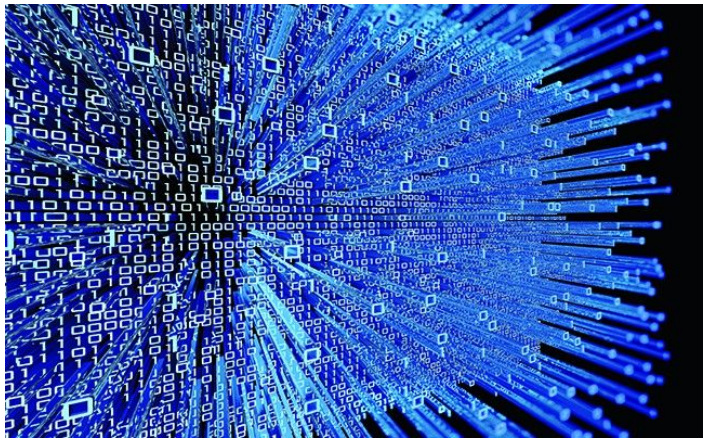
Two streams: melody and chords

Less parameters for DL model to learn

Encodes essential musical information while leaving room for personal creativity

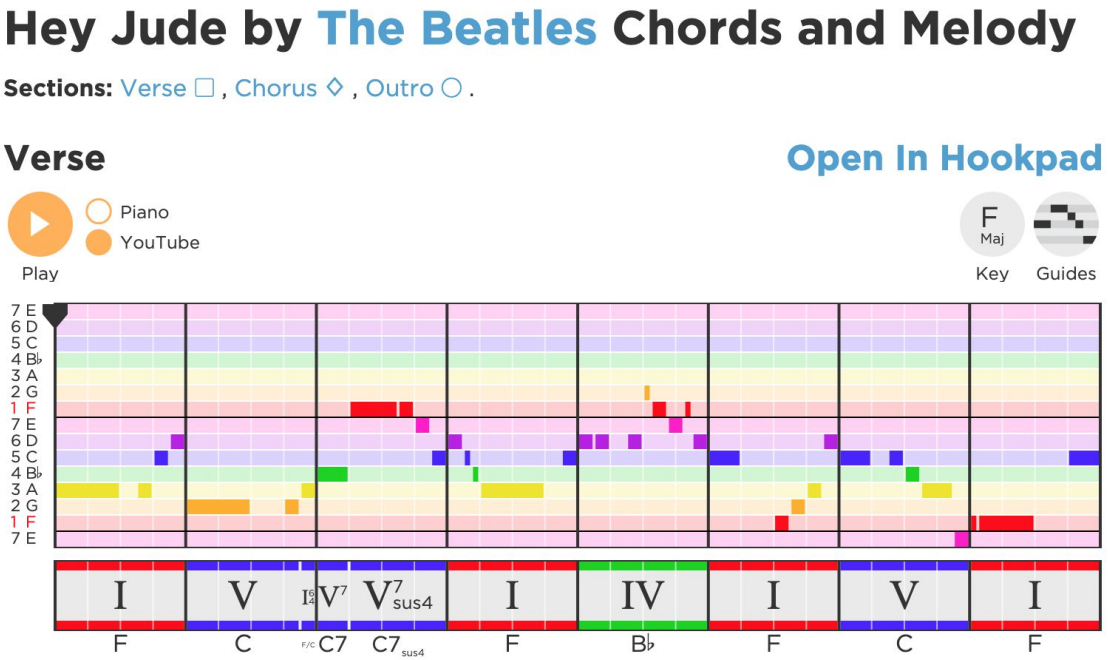


Dataset



HookTheory TheoryTab Database

More than 30,000 user-submitted snippets of music from a wide range of genres.



HookTheory TheoryTab Database

Advantages:

- Music is already pre-partitioned into sections like intro/verse/chorus/bridge.
- Music is already in lead-sheet-like form (i.e. melody and chords).
- Database is up-to-date with respect to U.S. pop.





HookTheory TheoryTab Database

Only hooks tagged under the “pop” genre were considered.

4,338 pop hooks in JSON format were obtained after web-scraping.

Imported into Google Colaboratory dev environment.

Description of Final Model

Preprocessing Steps

- Only keep hooks that stay in a single key, and standardize all keys to C.
- Only keep hooks in 4/4 time.
- Only keep hooks containing both melodies and chords.
- Only keep hooks with melody range < 2 octaves.
- **Ignore all rests (except for pickup rests)**
- **Simplify all chords to their root triadic forms.**
- **Modulo all note values to fit into a one-octave range (ouch!)**

Raw web-scraped dataset size: **4338**

Sanitized dataset size: **2917**

Dataset Encoding

A hook is a sequence of musical **events**.

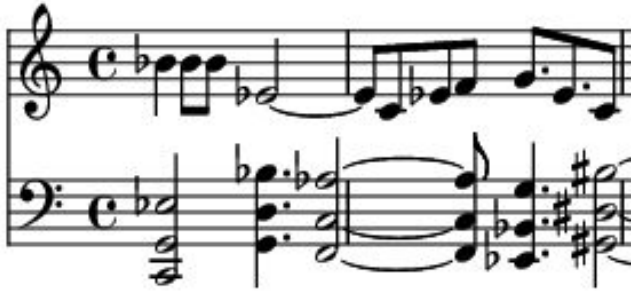
An **event** is the beginning of a note, the beginning of a chord, or time step 0.

Each event encodes four pieces of information: **note**, **chord**, **duration**, and **“position”**.

Special case:

1. If the next event is a new chord and the previous note is held, the note is *marked as “hold”.* (Unlike De Boom et al.)

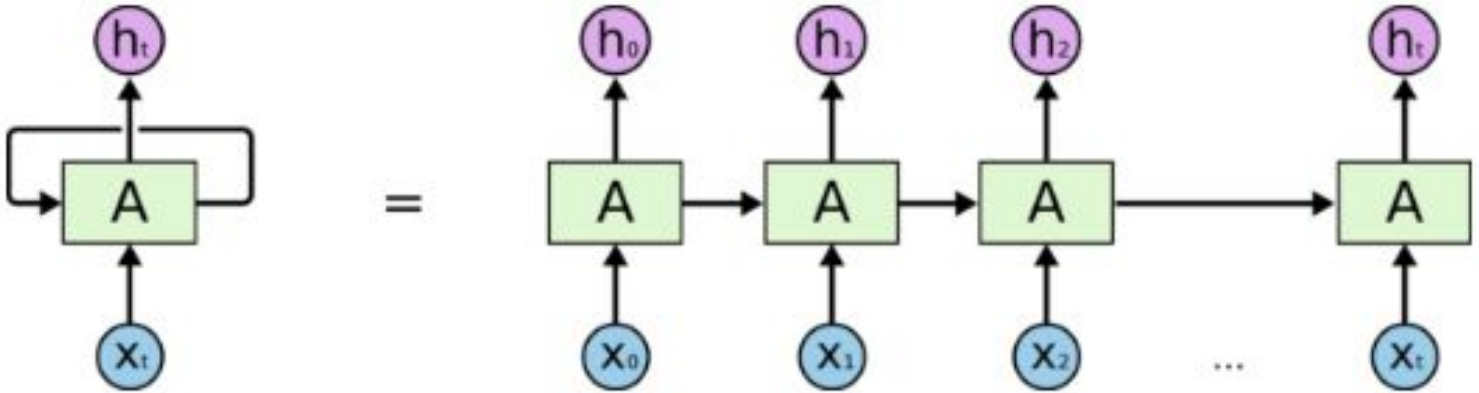
“I’m Coming Out” by Diana Ross



i	n_i	c_i	d_i	p_i
1	B♭4	C-min	4	0 (0000)
2	B♭4	C-min	2	4 (0100)
3	B♭4	C-min	2	6 (0110)
4	E♭4	G-min	6	8 (1000)
5	hold	F-min	4	14 (1110)
6	C4	F-min	2	2 (0010)
7	E♭4	F-min	2	4 (0100)
8	F4	F-min	2	6 (0110)
9	G4	E♭-maj	3	8 (1000)
10	E♭4	E♭-maj	3	11 (1011)
11	C4	A♭-maj	2	14 (1110)

Model Definition and Training

Recurrent Neural Networks (RNNs): NNs with **memory**



An unrolled recurrent neural network.

Model Definition and Training

Basic idea: given a input sequence of length *n*, predict the next item in the sequence.

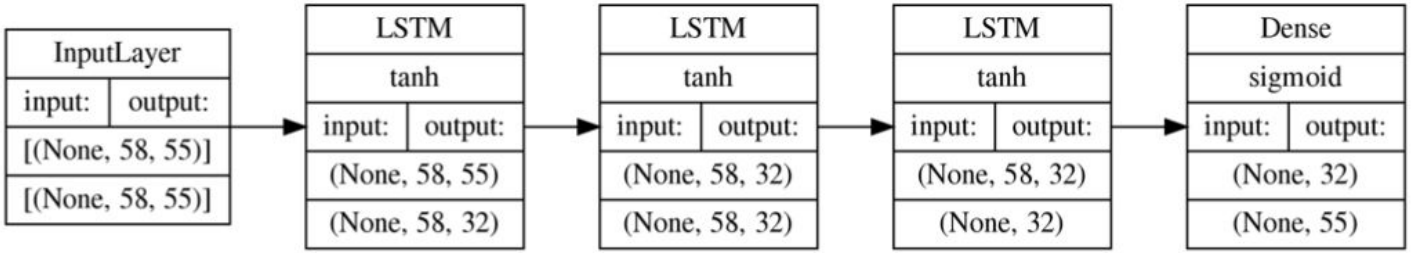
Sequences of events were split into **input** and **output** pairs.

Example: consider a hook with events *a, b, c, d, e, f, g.* Assuming ***n* = 3**:

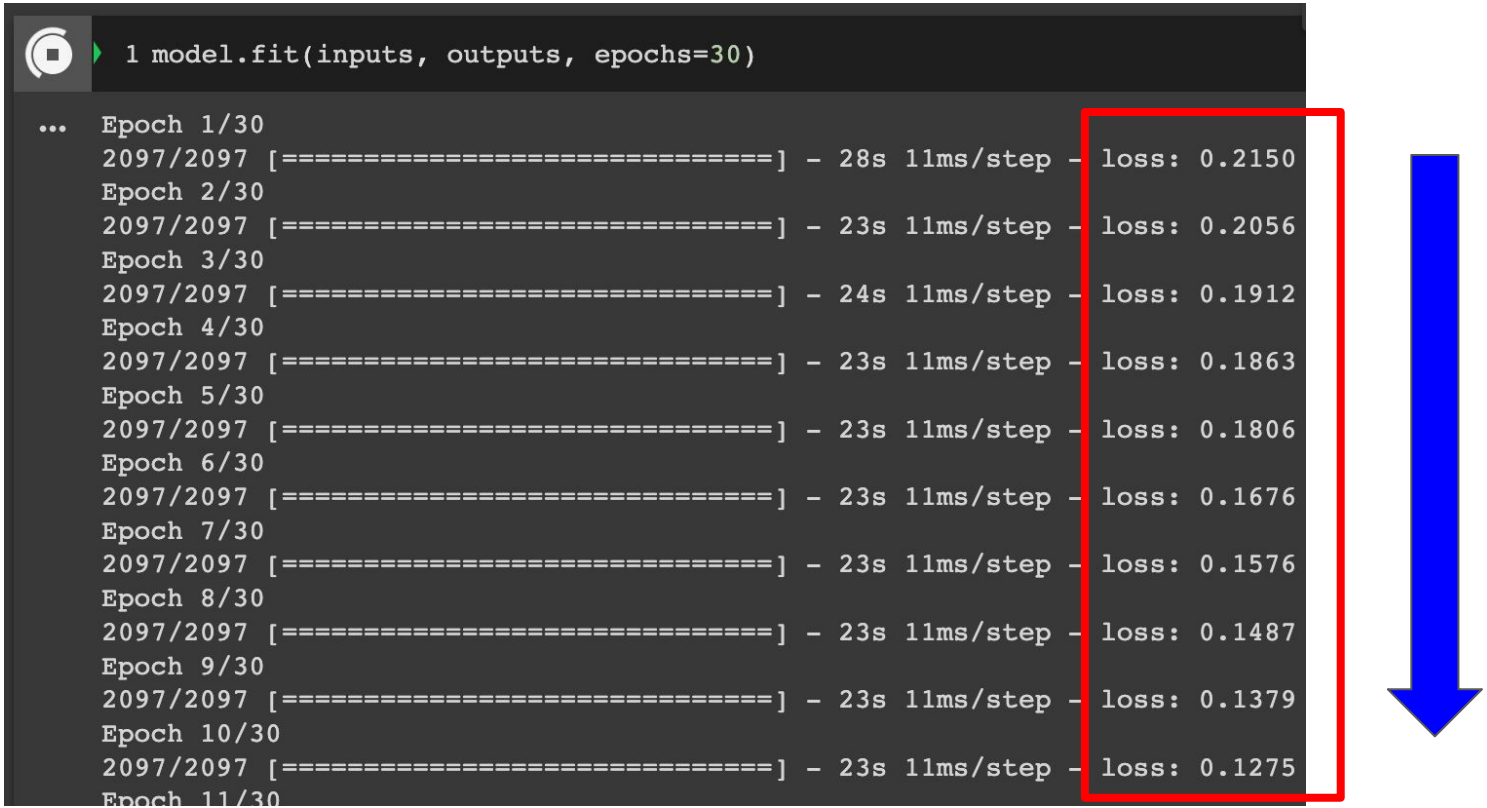
- Inputs: (*a*, *b*, *c*) (*b*, *c*, *d*) (*c*, *d*, *e*) (*d*, *e*, *f*)
- Outputs: *d* *e* *f* *g*

For this project: ***n* = 58** (somewhat arbitrary)

Model Definition and Training



Model Definition and Training



Music Generation Strategy

Simplified schematic, assuming *n* = 3:

```
Seed input: a, b, c
Prediction: δ
Input:      b, c, δ
Prediction: ε
Input:      c, δ, ε
Prediction: ζ
```

Generated hook:

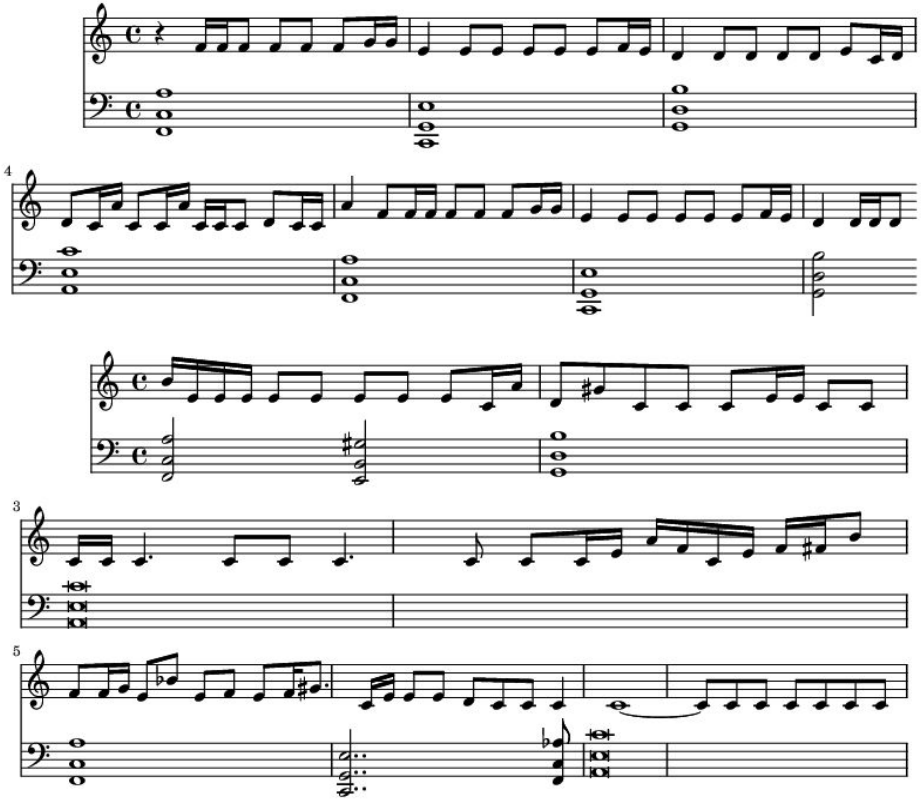
δ

ε

ζ

Etc. etc.

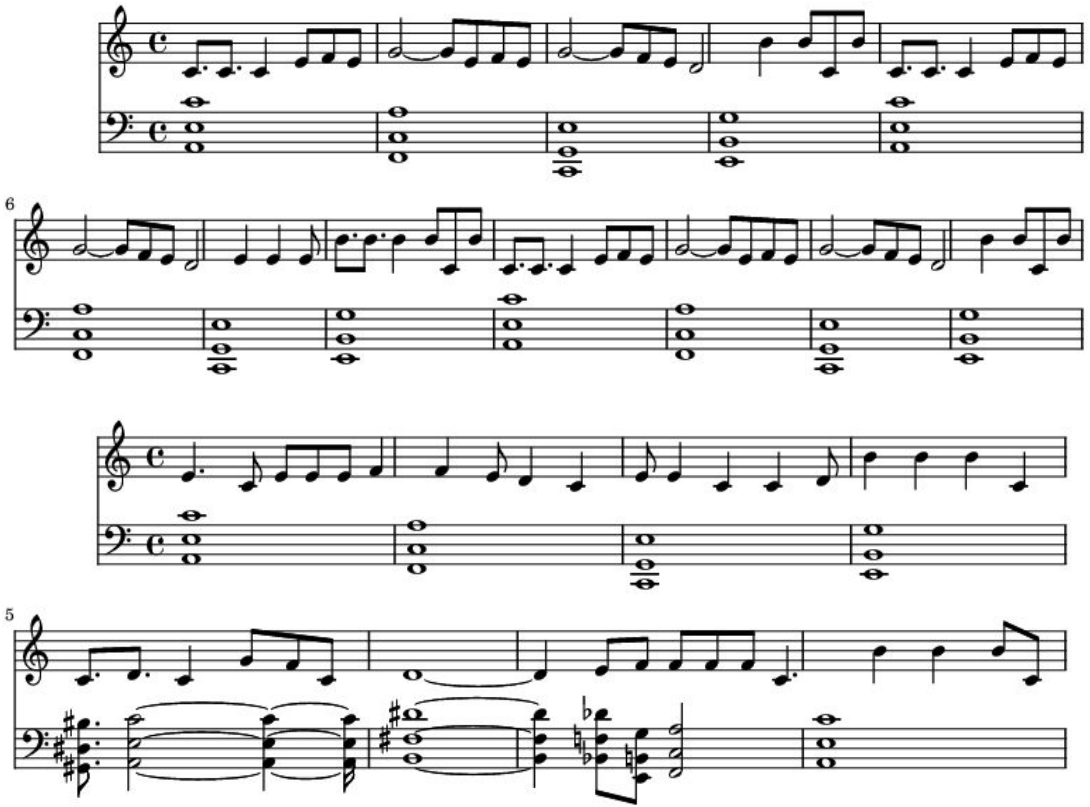
Generation Demo



Seed: “Umbrella” by Rihanna

Generated hook

Generation Demo



Seed: “A Sky Full Of Stars” by Coldplay

Generated hook

Analysis

Positive attributes

- The rhythms are realistic.
- Chord changes are particularly convincing (at least compared to previous models).
- The generated hooks appear to capture the harmonic and rhythmic essence of the seed without plagiarizing.

Negative attributes

- The output tends to get “lost” as the hook progresses
- The model still chooses funky/atonal notes and chords somewhat regularly.

Plagiarism Checker

Plagiarism Checker Algorithm

Remove durations and repeated notes in all melodies and encode as a sequence of integers.

- Example: Chorus of “Hey Soul Sister” by Train
- [7, 4, 2, 0, 7, 4, 2, 0, 2, 4, 0, 4, 0, 4, 2, 9, 0]

Use Python `SequenceMatcher` algorithm to compare melody in question against all melodies in the database.

Sort and print results.

What is plagiarism? Also why only melodies?

Song Example: “Will You Learn”

Using generated hook from “Sky Full Of Stars” seed (with some tweaks).

Produced and mixed with Logic.

Special thanks to **Claudia Humphrey ‘22** for her outstanding vocal talent (which vastly outshines my painfully mediocre songwriting abilities).

Discussion

Takeaways

When working with complex data like music, **encoding is the hardest part.**

Having **musical intuition** makes encoding easier.

Finding a **suitable dataset** can be challenging.

For music, the process of evaluating a model's output will necessarily be **more qualitative**.

Because neural networks are intrinsically **difficult to understand**, they are more **difficult to tweak**.

Future Work

Increase **dataset size**.

Try **different architectures/approaches**.

- Markov model, transformer architecture, etc.

Package into user-friendly **app interface**.

Add **user-selectable options**.

- Genre selection
- Type of hook (intro/verse/chorus/etc.)

Acknowledgements

Image Sources

https://www.collinsdictionary.com/us/dictionary/english/nutshell

https://time.com/4494496/tesla-mobileye-autopilot-safety/

https://openai.com/dall-e-2/

https://www.usatoday.com/story/tech/2018/01/11/facebook-newsfeed-big-change/1023331001/

https://www.amybergquist.com/blog/2008/10/02/pricing-a-home-price-per-square-foot/

https://www.takamine.com/g-series

https://www.istockphoto.com/photo/magician-or-illusionist-is-showing-magic-trick-blue-stage-light-in-background-gm912003020-251080485

https://research.google/teams/brain/magenta/

https://www.vecteezy.com/vector-art/2277576-narrowing-of-the-road-traffic-symbol-sign-isolate-on-white-background-vector-illustration

https://en.wikipedia.org/wiki/Billboard_Year-End_Hot_100_singles_of_1966

https://www.bu.edu/synapse/2011/11/27/earworms/

https://www.sheetmusicplus.com/title/autumn-leaves-digital-sheet-music/19426486

https://towardsdatascience.com/why-and-how-to-use-pandas-with-large-data-9594dda2ea4c

Image Sources (cont.)

https://www.hooktheory.com/

https://variety.com/2021/music/news/dua-lipa-dababy-homophobic-remarks-1235028796/

https://open.spotify.com/artist/6eUKZXaKkcviH0Ku9w2n3V

https://www.thecut.com/2021/04/bts-named-louis-vuittons-newest-house-ambassadors.html

https://en.wikipedia.org/wiki/Ariana_Grande

https://www.amazon.com/SOUR-Olivia-Rodrigo/dp/B091DXKLNT/ref=asc_df_B091DXKLNT/

https://www.redbubble.com/i/poster/bruce-lee-quotes-simplicity-is-the-key-to-brilliance-by-cocoCabot/69103550.LVTDI

https://keras.io/

https://www.heartland.org/news-opinion/news/the-future-is-bright-at-heartland

Thank you!