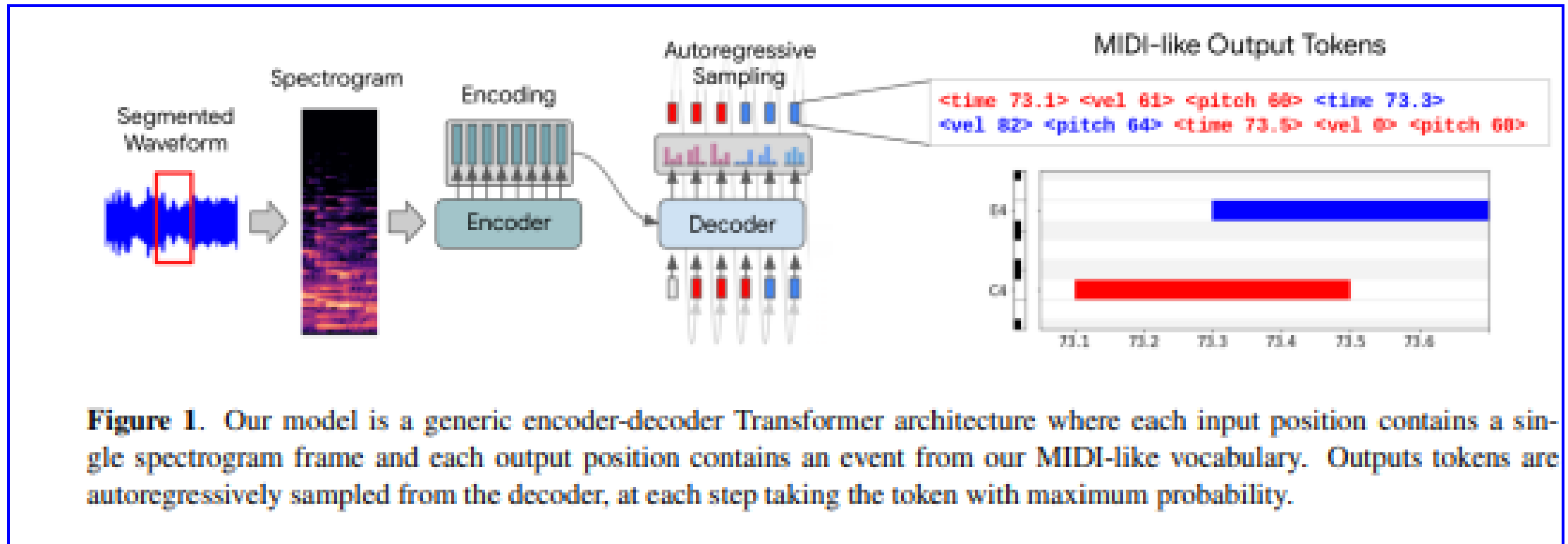


Audio-based AI approaches to music generation

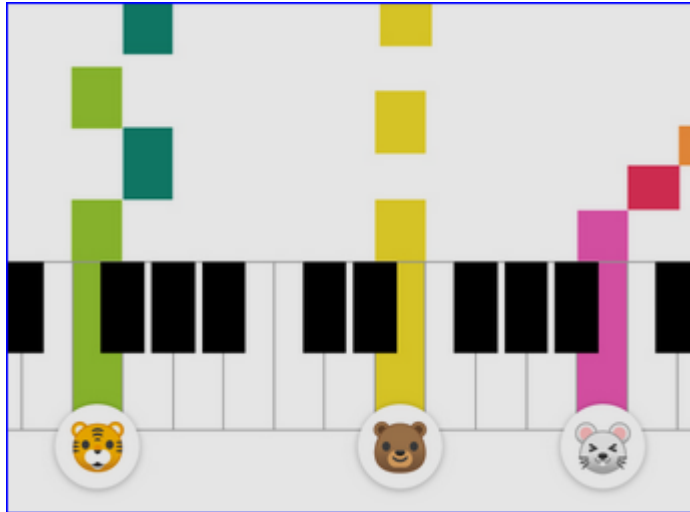
Google Magenta

- <https://magenta.tensorflow.org/blog>
- Aim: to provide tools for novices
- Piano transcription: Hawthorne et al. (ISMIR 2021):
- <https://archives.ismir.net/ismir2021/paper/000030.pdf>
- Compare various datasets

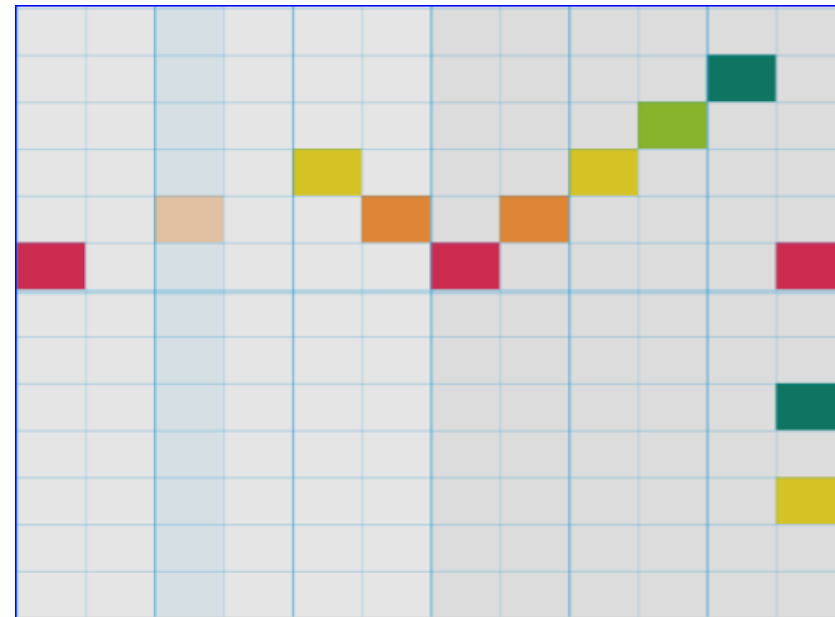


Other models: Chrome Music Lab

- Pedagogy, keyboard facilitation



SongMaker



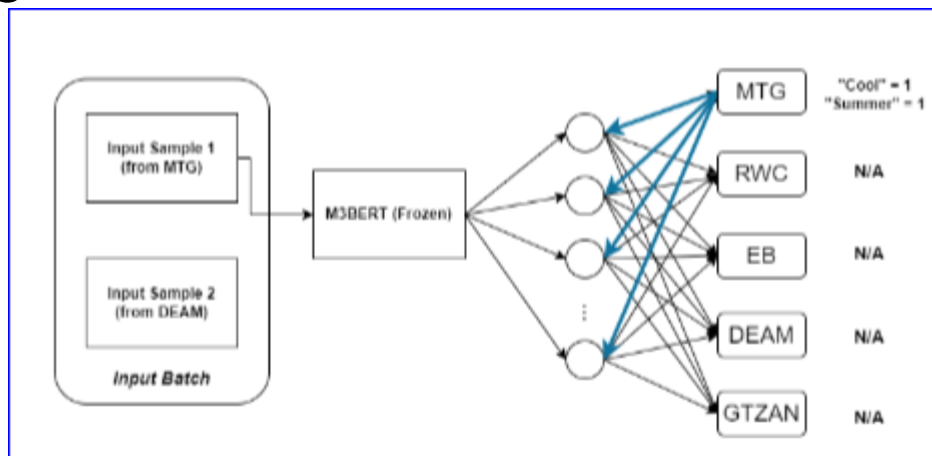
Hook theory/Hookpad

<https://www.hooktheory.com/>

The screenshot displays the Hook Theory software interface. At the top, there are playback controls (Play, Record, Loop, Click, Mixer, Preview) and settings for Meter (4 Beats), Key (C Major), Tempo (120 RPM), and Band. The main workspace shows a chord progression in C major: I (C), ii (dm), iii (em), IV (F), V (G), vi (am), and vii° (b°). The IV chord (F) is highlighted in green. Below the chord list is a piano roll with a grid for notes (B, A, G, F, E, D, C) and a timeline from measure 1 to 5. The progression is visualized as I (C), V (G), V (G), vi (am), and IV (F). The IV chord (F) is highlighted in green. On the right, the Chord Properties panel shows options for Type, Inversion, and Options.

Elaborate generation systems

- Timothy Greer et al., USC (2022): M3BERT
- Music generation via transformers; context aware



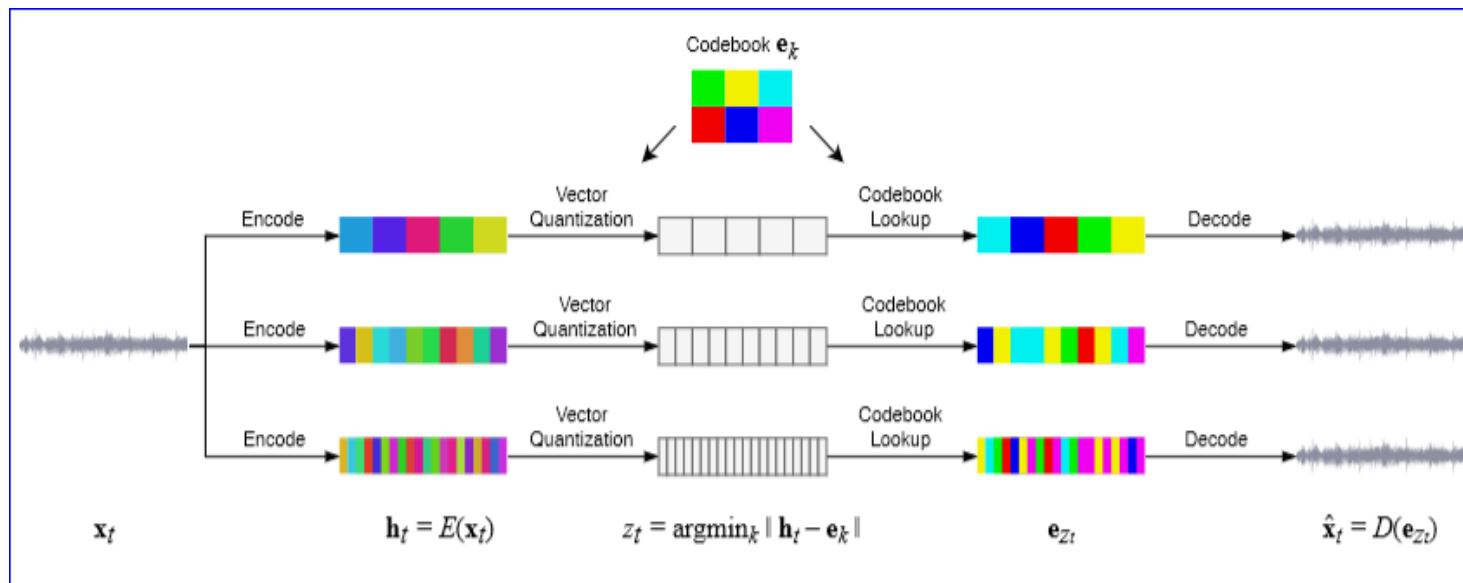
Cf. David Cope, *Experiments In Musical Intelligence*

https://www.researchgate.net/publication/363811441_Multi-modal_Multi-task_Music_BERT_A_Context-Aware_Music_Encoder_Based_on_Transformers/figures?lo=1

Generative singing: Jukebox

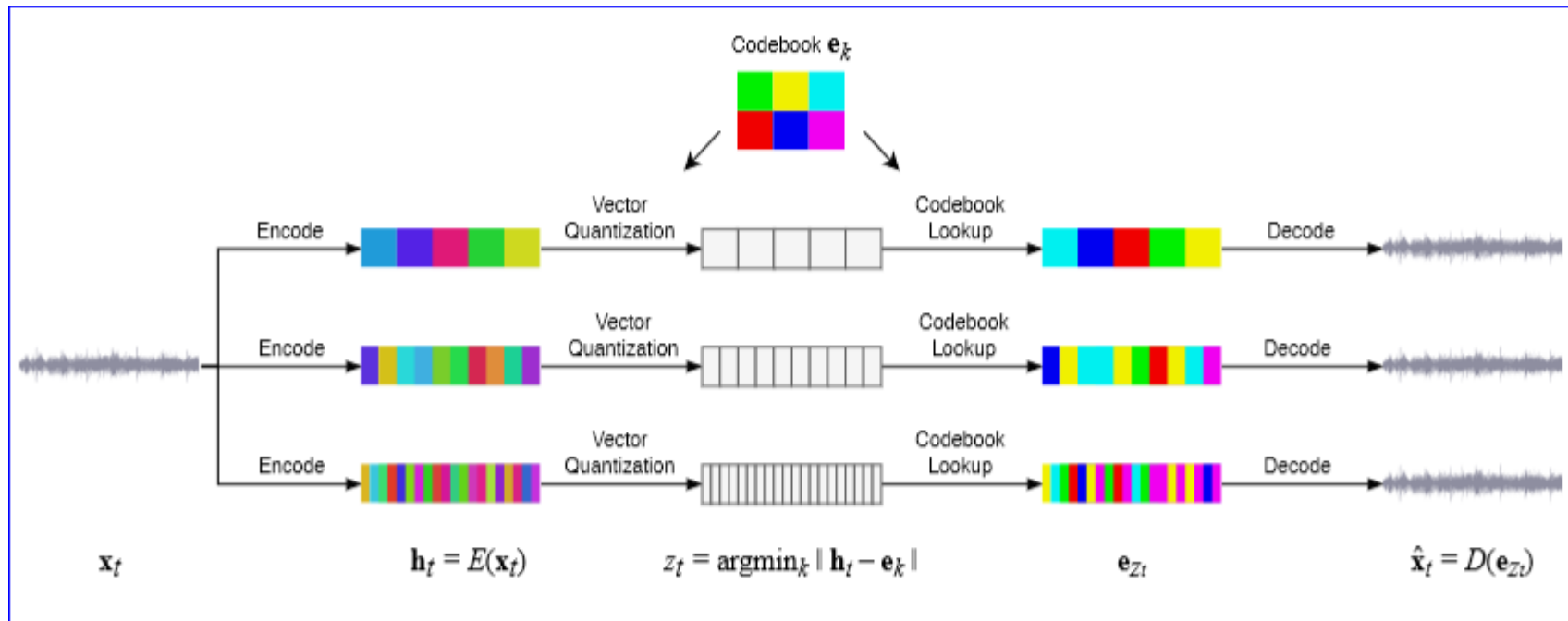
- [openAI \(2020\)--https://arxiv.org/pdf/2005.00341.pdf](https://arxiv.org/pdf/2005.00341.pdf)
- <https://openai.com/research/jukebox> (Dhariwal et al.)
- Using transformers with text-to-audio approach

Cf. Sapp, Keyscapes



Text-to-music generation via DALL-E2

- Tiers of temporal resolution (cf. Craig Sap, keyscapes)



MusicLM (Google)

- Emphasis on **audio quality** of generation (L-C Yang et al.)
- Preliminary writeup (MidiNet: ArXiv 2017)
 - Builds on Hawthorne et al, "Sequence to sequence" transcription"
- <https://arxiv.org/pdf/1703.10847.pdf>

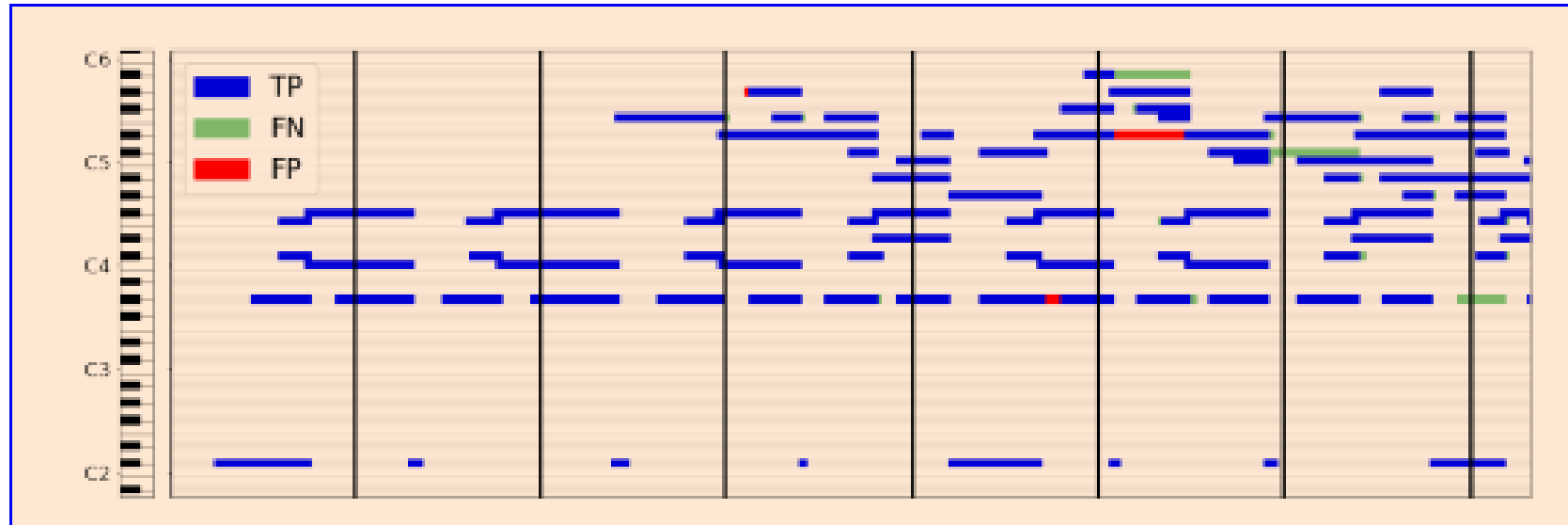
Hawthorne: Sequence-to-sequence studies

- Modeling music with NNs is difficult because it requires structures at many different time-scales.
- Developed the **MAESTRO** dataset to facilitate research.

<https://doi.org/10.48550/arXiv.1810.12247>

200 hours of paired audio and MIDI recordings from ten years of International Piano-e-Competition.

MAESTRO: Piano roll rendering of Chopin's Berceuse op 57



Harthorne et al., ISMIR 2021

- Piano transcription: "sequence-to-sequence" task
- Transcribing multiple instruments:
- cf. <https://magenta.tensorflow.org/transcription-with-transformers>

Failure to find AI superior to human composition among trained listeners (Tom Collins, Springer Verlag, April 2023)

- Collins article:
- <https://link.springer.com/article/10.1007/s10994-023-06309-w>
- Discusses MAESTRO and other recent datasets
- 50 well trained subjects
- First movement of classical-period string quartets (Haydn, Mozart, Beethoven) if in fast tempo
- Also used expressive MIDI data from Yamaha Disklaviers
- Collins uses MAIA Markov (non-deep learning) model

Criteria for evaluation

- Stylistic success
- Aesthetic pleasure
- Repetition
- Melody
- Harmony
- Rhythm

All criteria are rated higher for natural examples than for artificial examples.

Classifying users for Spotify recommendations

- Spotify: https://www.wsj.com/video/series/wsj-explains/how-spotify-knows-what-you-want-to-hear-next/E91EB935-C3EE-42FF-B41A-246614F8F1A1?mod=hp_listb_pos1 (WSJ, 16 April 2023)
- Explains layers of computation in building personal recommendations