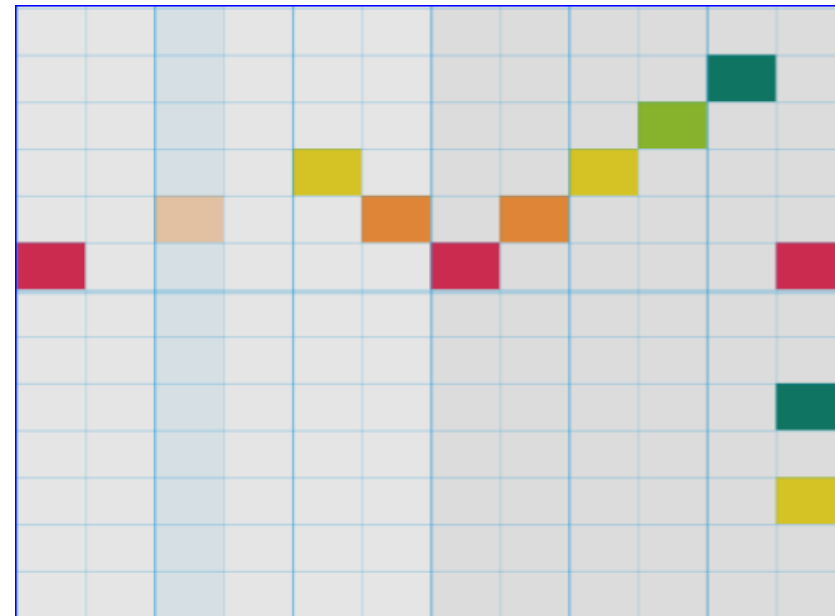
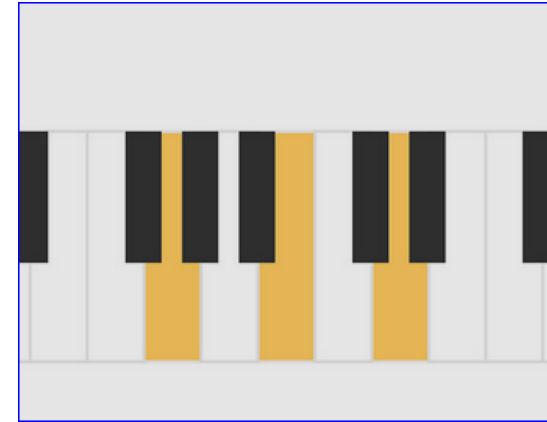
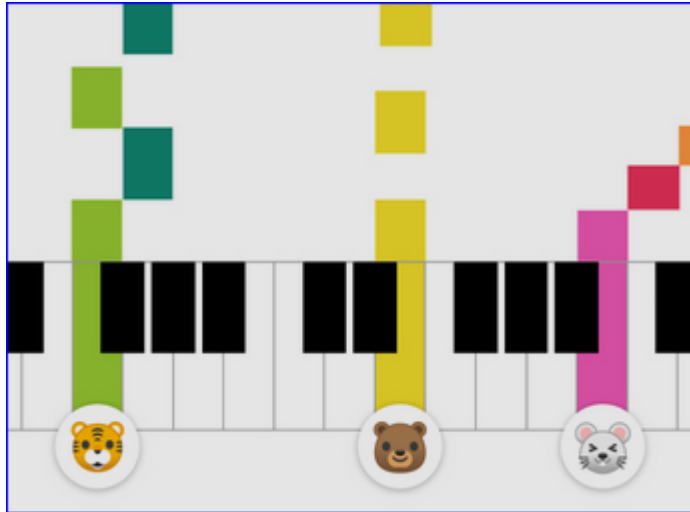


Audio-based AI approaches to music generation

Curtis Hawthorne, Anna Huang et al. (Google Magenta c.2017-2023).

Other models: Chrome Music Lab

- Pedagogy, keyboard facilitation



SongMaker

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

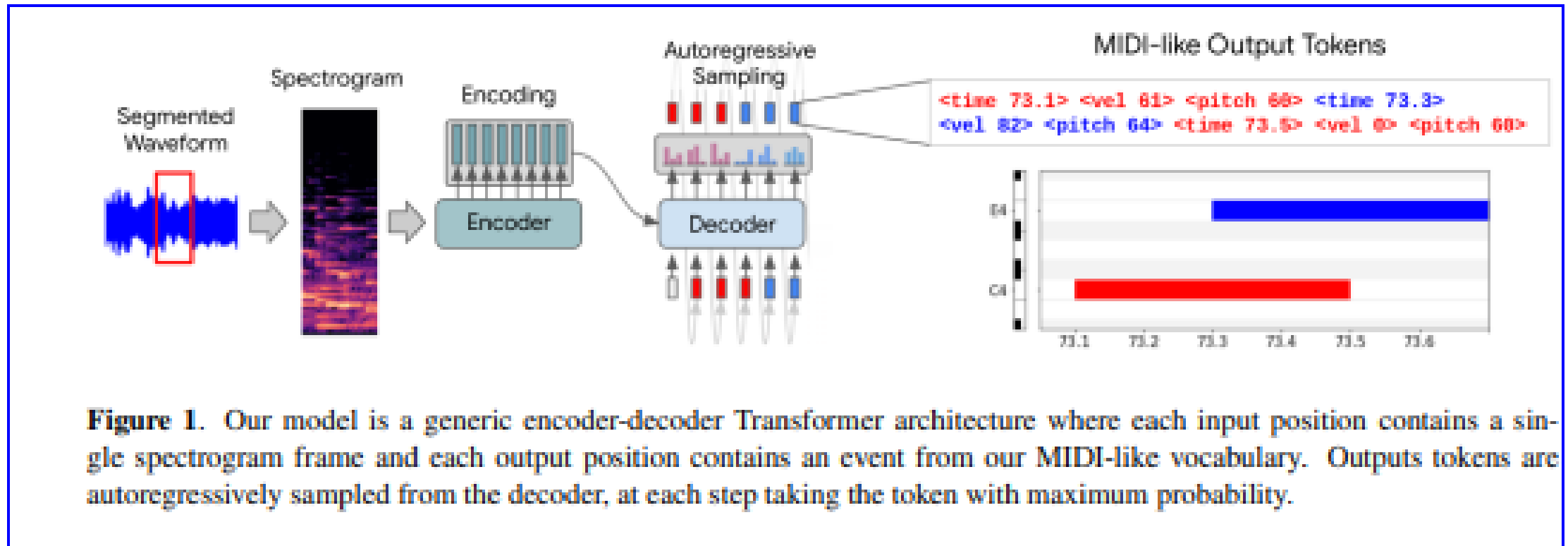
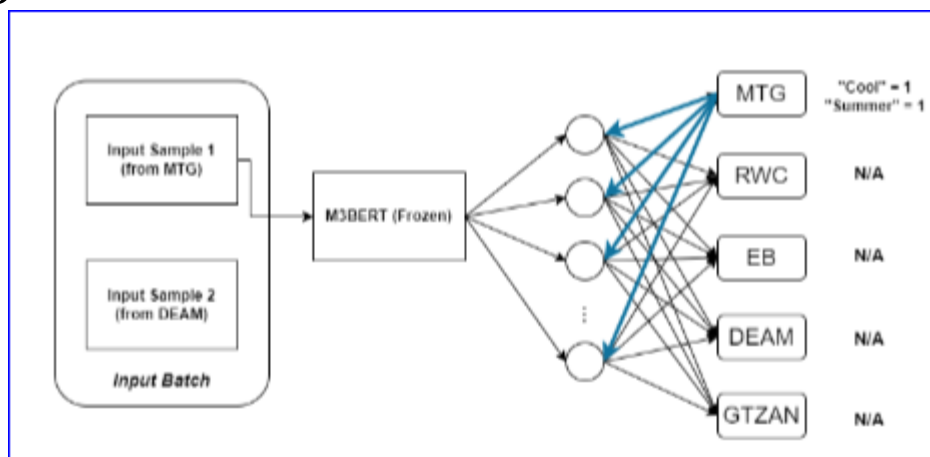


Figure 1. Our model is a generic encoder-decoder Transformer architecture where each input position contains a single spectrogram frame and each output position contains an event from our MIDI-like vocabulary. Outputs tokens are autoregressively sampled from the decoder, at each step taking the token with maximum probability.

Elaborate generation systems

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



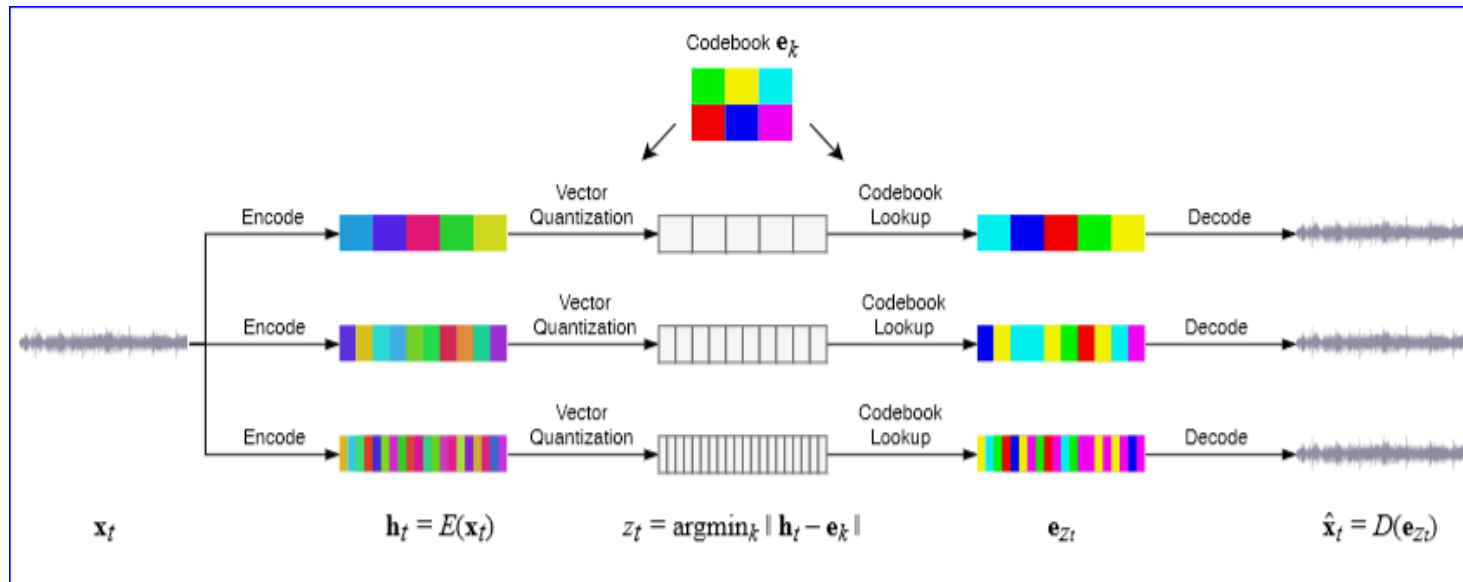
Cf. David Cope, *Experiments In Musical Intelligence*

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

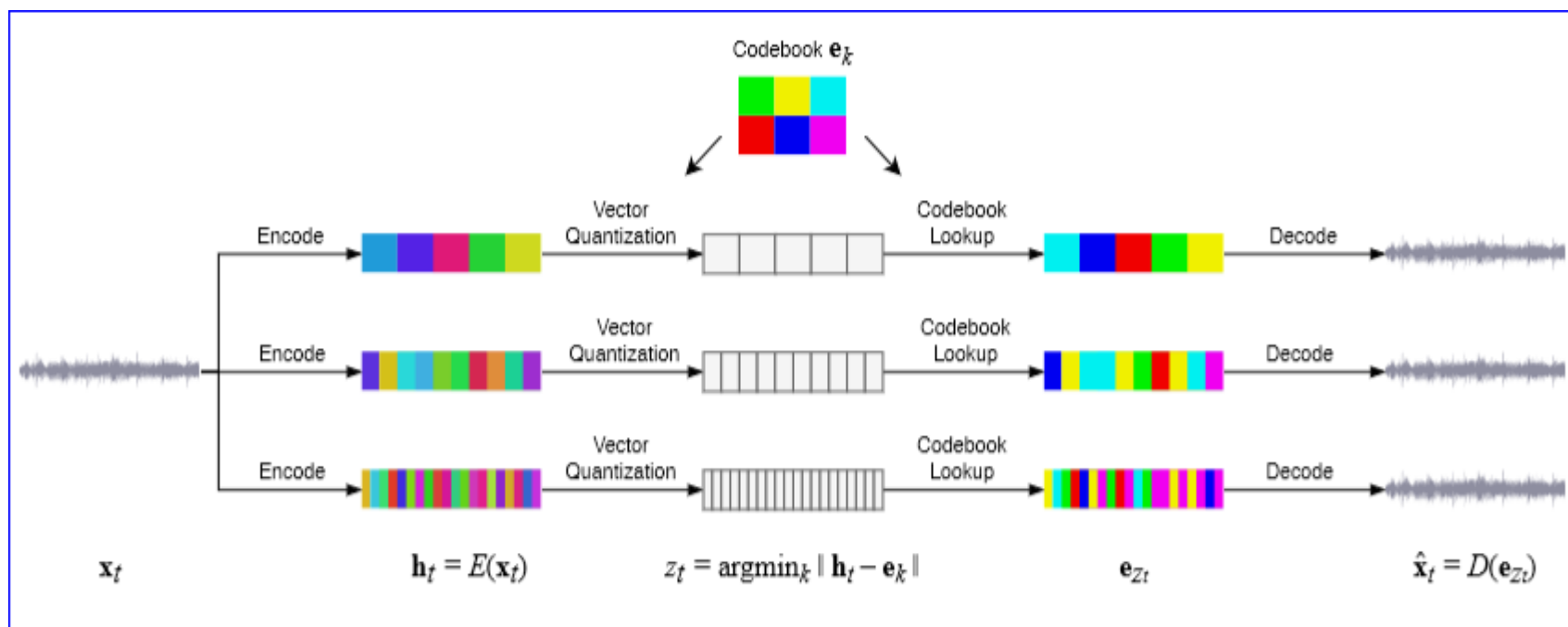
- [openAI \(2020\)--https://arxiv.org/pdf/2005.00341.pdf](https://arxiv.org/pdf/2005.00341.pdf)
- <https://openai.com/research/jukebox> (Dhariwal et al.)
- Using transformers with text-to-audio approach

Cf. Sapp, Keyscapes



Text-to-music generation via DALL-E2

- Tiers of temporal resolution (cf. Craig Sapp, 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]

Maestro Dataset (Curtis Hawthorne)

- Piano competition submission data (Wisconsin, Disklavier submissions)
- Vivian Shaylin Chen (SU, 2021): same datasets, different procedures—What performance traits produce best contest judgments?

Vivian Chen: Chopin “Waterfall” Etude contestants

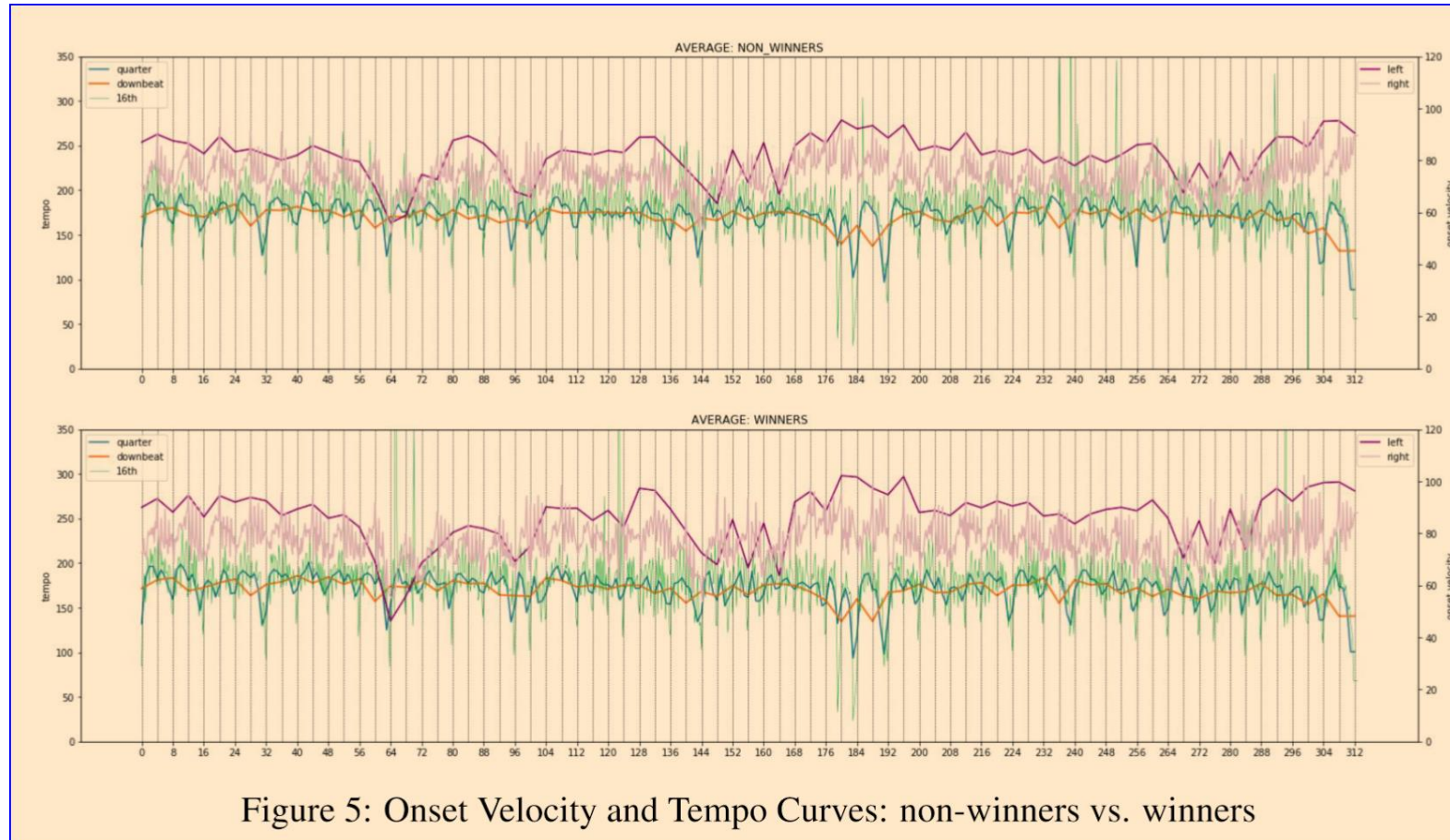
Ardeeva vs Sun (No. 2): treatment of sixteenth notes with respect to evenness, consistency

The image displays a musical score for Chopin's "Waterfall" Etude, comparing two performances of the 16th note placement. The score is presented in two systems, each with a treble and bass clef. The notes are color-coded to highlight differences in placement and timing. Red notes indicate a specific placement, while blue notes indicate a different placement. Dashed lines with the number '8' above them indicate eighth-note groupings. The score is numbered 5 and 9 at the beginning of the systems.

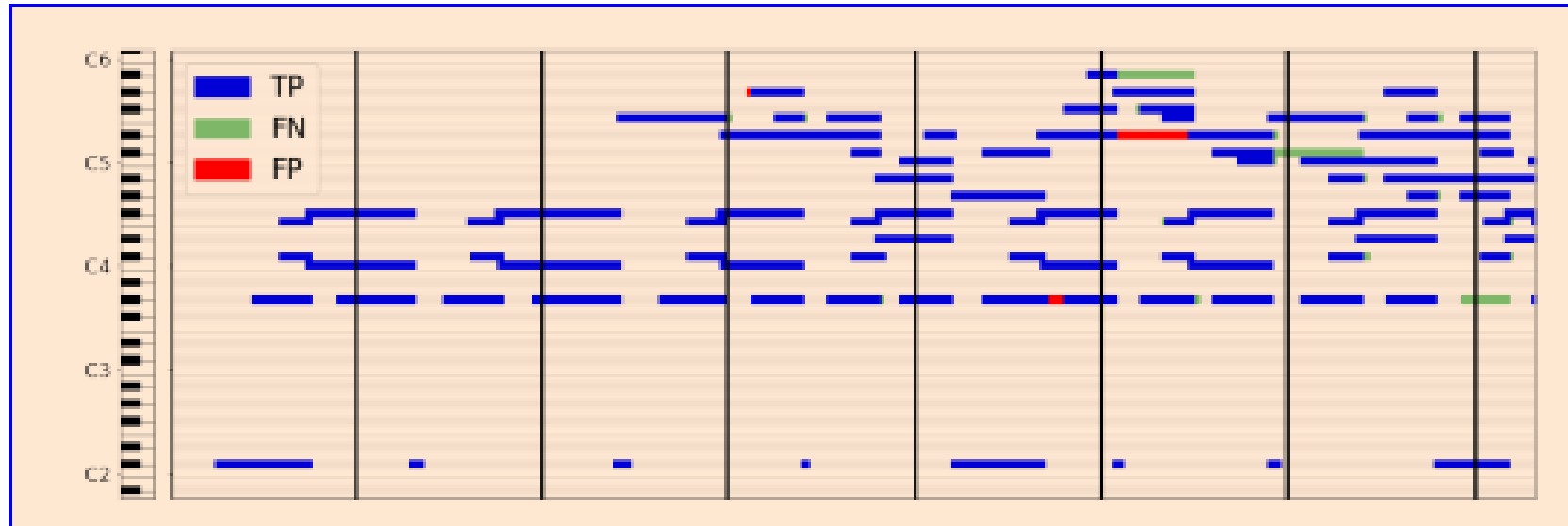
Figure 7: 16th note placement by color, comparing Avdeeva with Sun (2nd place)

Chen: Onset velocity and tempo curves (2021)

Collective winners, non-winners



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



<https://chopinscores.org/pl/partytura/283>

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

Collins's Criteria for evaluation

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

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

Tom Collins et al (UK, 2011...)

- Assessments of algorithmic composition methods
- Computational models of musical style
- Databases offering clusters of similar works
 - Haydn/Mozart string quartets [qq.themefinder.org] cf. Kempfert
 - Chopin Mazurkas [N. Cook mazurka recordings]
- Procedural approaches [IEEE]
 - Sequential
 - Structural

Tom Collins et al York U. (2011...)

- 2023, ML, col 1789:
- “The **lack of full source code** and description of how the model works has attracted criticism and called the EMI project into question “ (Wiggins, 2008; Collins et al., 2016).
- Collins’ description of the process focuses on “**nesting a Markov generator in another form that inherits a long-term repetitive structure (MAIA)**”.

Collins

- <https://www.cambridge.org/core/journals/ai-edam/article/developing-and-evaluating-computational-models-of-musical-style/2D13038AEC3BB894F1345C63F74F6CF4>

Emmy vs Chopin

1
6
11
16
21
25

Score (a) shows the original notation for Emmy's version. It consists of six systems of music, each with a treble and bass clef staff. The key signature is three flats (B-flat, E-flat, A-flat) and the time signature is 3/4. The piece begins with a first-measure rest in both staves. The melody in the treble staff is characterized by eighth-note patterns and slurs. The bass staff provides harmonic support with chords and single notes.

(a)

1
6
11
16
21
25

Score (b) shows the original notation for Chopin's version. It follows the same structure as score (a), with six systems of music in the same key and time signature. The notation is more fluid and expressive, with many slurs and grace notes. The first-measure rest is present in both staves. The melodic lines in the treble staff are more intricate and varied in rhythm compared to Emmy's version.

(b)

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