Probing for Advanced Music Theory Concepts in Generative Music Models

Derek Kwan and Patrick J. Donnelly

Oregon State University

Introduction

We constructed the **SynTheory-Plus** dataset centered around five advanced music theory concepts to probe OpenAI's Jukebox and Meta's MusicGen-small, -medium, and -large generative music models as well as Meta's EnCodec audio encoder. This work builds upon Wei et al.'s [1] SynTheory dataset centered around the basic music theory concepts of notes, scales, intervals, triad chords, triad chord progressions, tempi, and time signatures used to probe the same generative models.

Probing trains classifiers on transformer layer outputs to discriminate between properties of a dataset concept. High evaluation accuracies indicate that the concept is encoded in a model's latent space. This encoding is required for concept property generation. We focus on demonstrating the encoding of concepts but leave their generation for future work.

Methods

We first **create our concept datasets** by programmatically generating MIDI files through the Python `mido` library. Following Wei et al. [1], these MIDI files are rendered into WAV files using `RustySynth` and the `TimGM6mb` soundfont.

Table 1. SynTheory-Plus Dataset Overview

Dataset	Examples	Classes	Pitch-Centric	Balanced
Polyrhythms	30,600	34	No	Yes
Dynamics	41,580	7	No	No
Seventh Chords	35,328	8	Yes	Yes
Mode Mixture	19,872	2	Yes	Yes
Secondary Dominants	26,496	3	Yes	Yes

To **extract the WAV files' corresponding decoder layer outputs** from Jukebox (JB), we use the `jukemirlib` Python library. To extract outputs from MusicGen-small (MGS), -medium (MGM), -large (MGL), and EnCodec (EnC), we follow Wei et al. [1] and use the Hugging Face API.

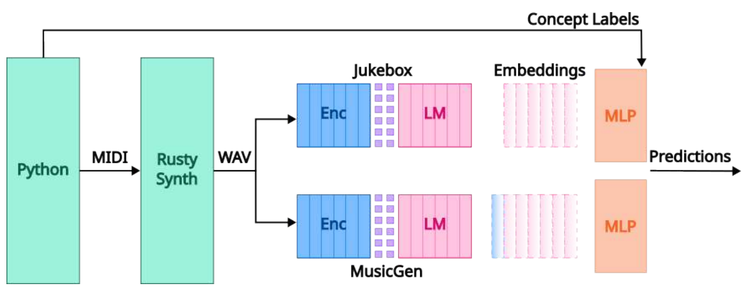


Figure 1. Workflow of our experiments.

In our **probing experiments**, we train linear layer and one-hidden-layer MLP classifiers for each transformer layer in each dataset/model combination using a 70%/15%/15% training/validation/evaluation class-stratified data split.

huggingface.co/datasets/derekxkwan/syntheory_plus

SynTheory-Plus Dataset

SynTheory-Plus is centered around five concepts related to rhythm, loudness, and harmony.

Our rhythmic dataset consists of **polyrhythms** (PR) defined by 34 different rate ratios $r_i : r_j$ (where $r_i, r_j = 2, \ldots, 11$ and are coprime) realized in 4/4 time at 60, 120, and 180 BPM.

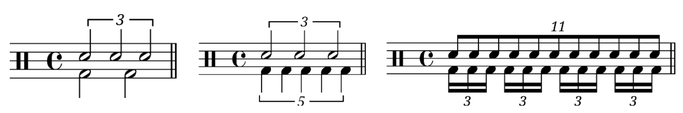


Figure 2. Polyrhythm ratios 2:3, 3:5, and 11:12.

Our loudness dataset consists of seven different **dynamic** (Dyn.) expression patterns (flat, subito piano, subito forte, crescendo, decrescendo, hairpin, reverse hairpin) realized over a stream of notes with quarter-note subdivisions 2 through 8 in 4/4 at 60 BPM. Where applicable, we modify the inflection point of dynamic direction.

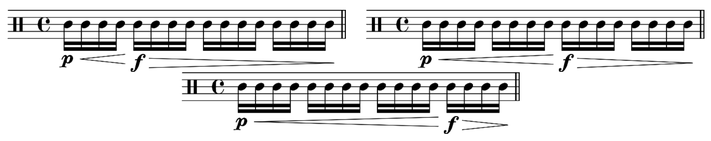


Figure 3. "Hairpin" with inflection points at beats 2-4.

The **seventh chords** (Ch7.) dataset consists of the eight chord qualities major 7^{th}, minor 7^{th}, major-minor 7^{th}, minor-major 7^{th}, half-diminished 7^{th}, fully diminished 7^{th}, augmented major 7^{th}, and augmented 7^{th}. Each chord is realized as four quarter-notes in one 4/4 bar with closed voicing at 60 BPM.

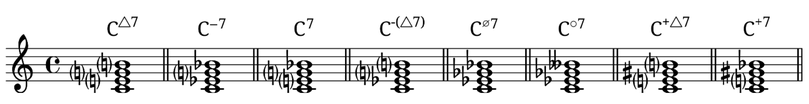


Figure 4. Eight seventh chord qualities.

Our two chord progression datasets focus on mode mixture and secondary dominants. The **mode mixture** (MM) dataset consists of select major triad chord progressions from SynTheory by Wei et al. [1] but paired with new counterparts where iv replaces IV and ♭VI replaces vi. The **secondary dominants** (SD) dataset consists of diatonic circle-of-fifths seventh chord progressions and variants where one chord is replaced with either the secondary dominant of the next or the tritone substitution of the secondary dominant. The tasks are to detect mode mixture and to detect the use of diatonic, secondary dominant, or tritone substitution variants.

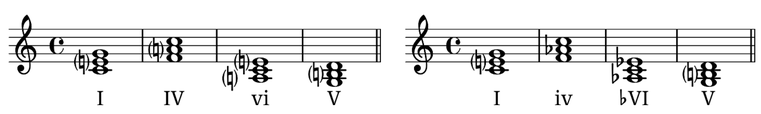


Figure 5. Regular and mode mixture chord progression.

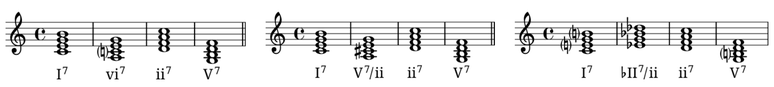


Figure 6. Diatonic, secondary dominant, and tritone substitution seventh chord progressions.

Probing Results

In addition to probing generative models, we conducted baseline experiments on audio features (where "All" is the three baseline features concatenated). Accuracy scores (balanced for Dyn.) correspond to evaluation on the transformer layer yielding the highest validation scores found through exhaustive search. Colors indicate $\geq .90$, $\geq 2 \cdot$chance, and $\geq$ chance.

Table 2. MLP Probe Results

Model	PR	Dyn.	Ch7	MM	SD
JB	1.0000	0.9990	0.9958	0.9980	0.9814
MGL	0.9998	0.9684	0.9764	0.9919	0.7560
MGM	0.9976	0.9527	0.9751	0.9802	0.6574
MGS	0.9935	0.9432	0.9724	0.9936	0.7137
EnC	0.7155	0.5376	0.6738	0.8916	0.5279
Mel.	0.4871	0.6572	0.8399	0.9671	0.6203
MFCC	0.7240	0.7720	0.5040	0.8100	0.5783
Chroma	0.3695	0.3933	0.9560	0.9973	0.8708
All	0.7608	0.8600	0.9322	0.9879	0.7825

Table 3. Linear Probe Results

Model	PR	Dyn.	Ch7	MM	SD
JB	1.0000	0.9992	0.9564	0.9970	0.9227
MGL	0.9847	0.9936	0.6863	0.8916	0.6684
MGM	0.9503	0.9825	0.5649	0.8701	0.6735
MGS	0.8481	0.9530	0.5915	0.7909	0.6153
EnC	0.2617	0.3850	0.2114	0.5532	0.3784
Mel.	0.2928	0.6484	0.2914	0.6600	0.4031
MFCC	0.1776	0.5964	0.2510	0.5955	0.4230
Chroma	0.1251	0.2258	0.1665	0.6945	0.5846
All	0.4257	0.7879	0.3656	0.7966	0.6365

Jukebox's probed decoder consists of 72 layers while MusicGen-small's decoder consists of 24 layers. MusicGen-medium's and -large's decoders both have 48 layers.

Table 4. Best MLP Layer Indices (1-indexed)

Model	PR	Dyn.	Ch7.	MM	SD
JB	54	47	58	56	58
MGL	33	33	43	44	41
MGM	34	31	46	33	33
MGS	10	16	14	11	13

Conclusion

We have constructed the **SynTheory-Plus** dataset centered on the advanced music theory concepts of polyrhythms, dynamics, seventh chords, mode mixture, and secondary dominants and have shown that these concepts are **encoded** within the Jukebox, MusicGen-small, -medium, and -large generative models' latent spaces as well as by the EnCodec encoder. Furthermore, we have shown that these latent space encodings and concept properties exhibit strong **linear relationships**.

References

[1] Megan Wei, Michael Freeman, Chris Donahue, and Chen Sun. **Do music generation models encode music theory?** In *Proceedings of the 25th International Society for Music Information Retrieval Conference*, pages 680–687. ISMIR, November 2024.

Oregon State University

soundbendor lab