What is Missing in Deep Music Generation?
A Study of Repetition and Structure in Popular Music

Shuqi Dai, Huiran Yu, Roger B. Dannenberg

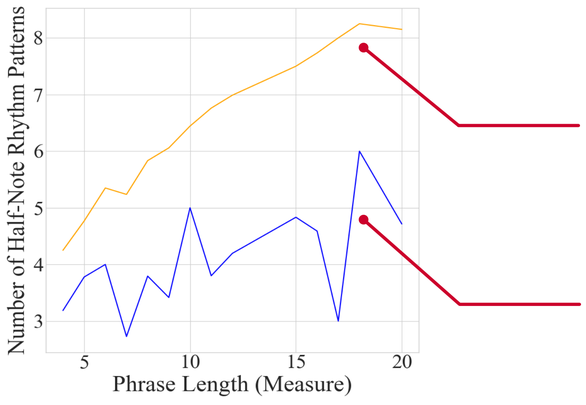
Contact:
shuqid@cs.cmu.edu

Key Findings

- **Repetition and structure** exist at multiple **hierarchical levels** in music, interacting with **rhythm, melody, harmony** and **predictability**
- Each song has **limited** and **specific vocabulary** different from entire dataset distribution
- **Repetition** is **not random** but follows a general **plan** over the course of a song
- Deep model generated music **lacks the structure** of real songs revealed by **objective measurement**

Study of Repetition and Structure

- **Multi-level Repetition Structure**
  - **Repetition reduces** number of **patterns** within phrases

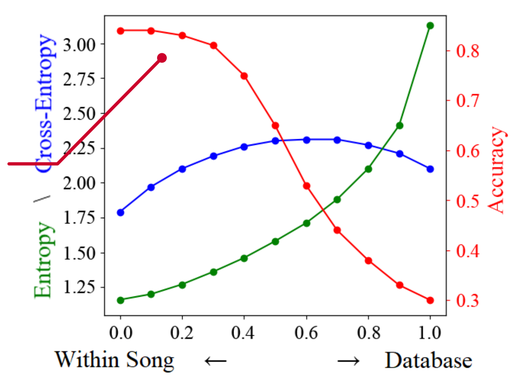


Number of patterns in phrase sampled from the **distribution of the database**

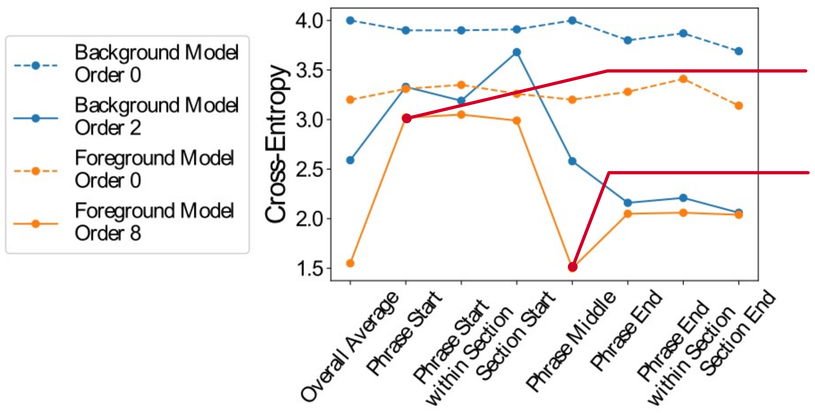
Number of patterns in **real** phrases

- **Song-Specific Vocabulary**
  - **Repetitions within song** help to make better predictions

Model trained with **song-specific** phrases reaches **highest prediction accuracy.**



- **Structural Influence**
  - The cross-entropy are **different** at different **structural positions.**



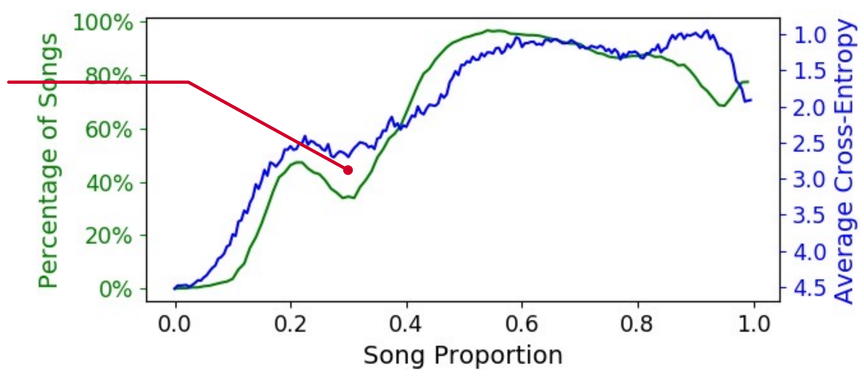
The predictability of notes are **different** at **start** and **middle** of the phrase.

- **Repetition Overtime**

The **schema** of repetitions and surprises over the song.

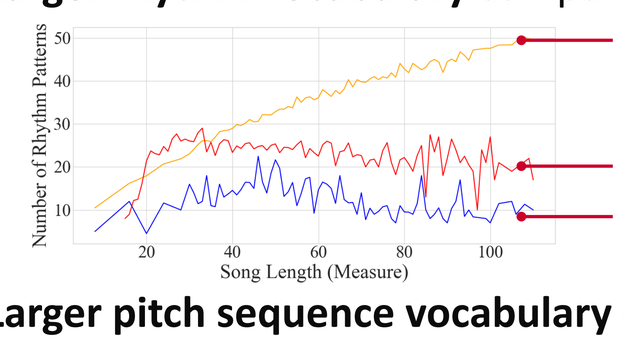
(To show the correspondence, we flipped the vertical cross-entropy axis)

New materials introduced



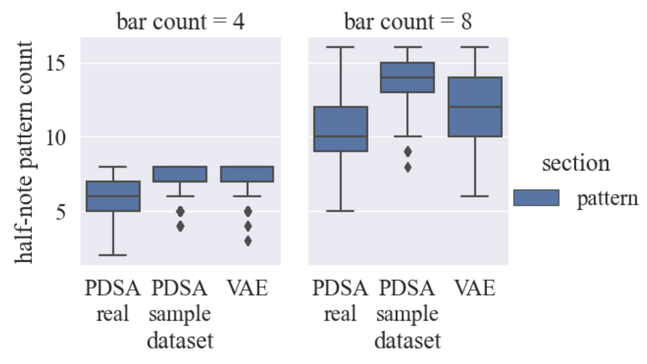
What's Missing in Deep Networks

- **Larger rhythm vocabulary** compared to real songs

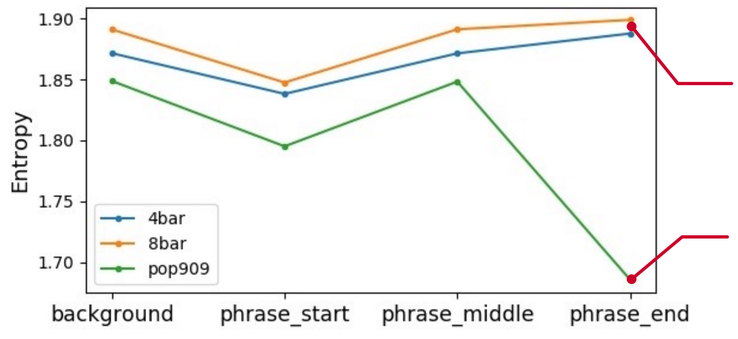


- **Larger pitch sequence vocabulary** compared to real songs

Real songs have **smaller** pitch sequence vocabulary than VAE model



- Positions in structure **do not affect** the distribution entropy



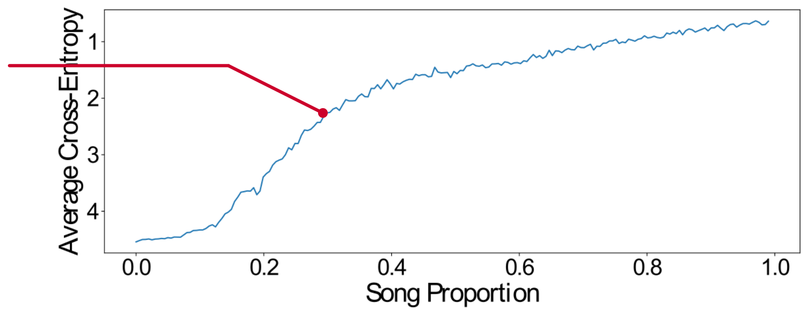
Deep network phrase ending pitch has **high** entropy

Real music phrase ending pitch has **lower** entropy

- **No** over-time cross-entropy **schema**

(We flipped the vertical cross-entropy axis)

No significant new materials introduced



Discussion and New Directions

- We need to learn how songs strategically **diverge from background** or stylistic norms to create interest, surprise, and individuality.
- We can compare generated music to real music using data-driven **measures of structure, repetition and entropy**.
- **Repetition and hierarchical structure are not restricted to pop music.** Based on the results, these studies can be used to reveal repetition and structure information in Classical and other types of music.

Conclusions

- Structure, repetition, pitch, rhythm, harmony and entropy are all **strongly connected and interdependent**, revealed objectively by our **data-driven** approach.
- Songs and phrases gain **individuality** through **more repetition and smaller vocabulary**, which are not a reflection of the general background statistics from a collection of songs.
- There are **clear differences** between metrics of real songs and those of many music generation systems, suggesting **important gaps to fill** for new research.