

Pop Music Generation with Controllable Phrase Lengths



Daiki Naruse¹, Tomoyuki Takahata¹, Yusuke Mukuta^{1,2}, Tatsuya Harada^{1,2} ¹ The University of Tokyo, ² RIKEN

Abstract

Goal

To control phrase lengths with a sequential generation policy

Input: Phrase configuration (label and length of each phrase)

Output: A piece with designated phrase lengths { •

- Phrase switching at designated position
- Natural closure at the end of outro

Phrase and Bar Countdown Events

We added Phrase and Bar Countrown events to the existing event-based music representations (REMI [1] and CP [2]) and showed that both events are necessary

Method

User Input

We extended autoregressive generation to reflect user input

[1] Huang+, ACMMM 2020 [2] Hsiao+, AAAI 2021

i4 A8 B8 o4

Introduction

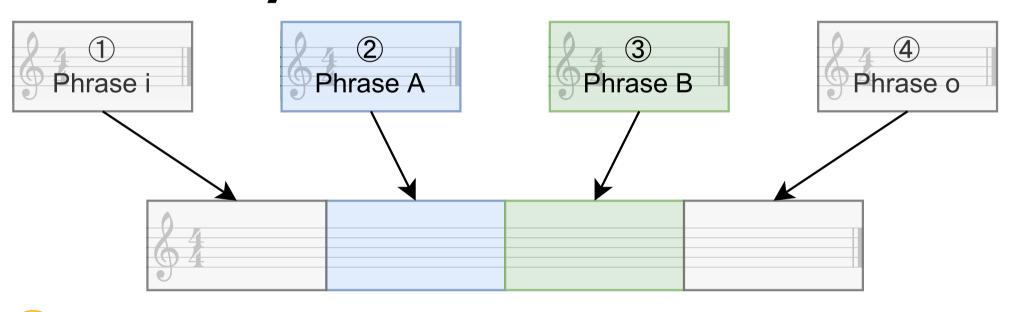
Contribution

Motivation

- The control of phrase attributes is important
- We focused on phrase lengths

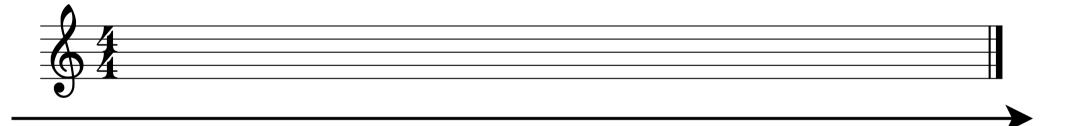
Two Strategies

Phrase-by-Phrase Generation



- Phrase lengths can be designated [3]
- No guarantee of natural transitions

Sequential Generation



- Transitions are natural [1, 2]
- No mechanism of phrase switching
- → We improved sequential generation to control phrase lengths
- [1] Huang+, ACMMM 2020 [2] Hsiao+, AAAI 2021 [3] Dai+, ISMIR 2021

Generated Pieces

Samples of the generated pieces are available at



Phrase Start Bar Countdown PHRASE (Ph): Which phrase a bar belongs to BAR COUNTDOWN (BC): The number of bars remaining in a phrase **Examples of Event-Token Sequence** CP + Ph&BC Bar Countdown REMI + Ph&BC Z Z Z B **Generation Process — How to Reflect User Input? Event Token Sequence** MIDI **Start Event Tokens** Bar Bar Neural Sequence Phrase (Start) ואירנונו 🎖 אינוע נועיו Phrase (Start) Bar Countdown (1) Model Bar Countdown (1)

Phrase (i) Bar Countdown (4) 1 Start event tokens are input to the model to start generation

Bar

Phrase (Start)

Bar Countdown (1)

Bar

- 2 Next event tokens are predicted by the model
- 3 Phrase and Bar Countdown are added based on user input

Experiments

Listening Tests

Test 1: Phrase switching at designated position

Subjects identified 3 phrase boundaries

Test 2: Natural closure at end of outro

natural or abrupt	
Results	

Subjects answered	whether	closure	was
natural or abrupt			

Method	Test 1 Phrase Switching	Test 2 Natural Closure	
	- I mase 5 wheming		
POP909	0.778	3.67	
REMI + Ph&BC (ours)	0.633	3.00	
REMI + BC	0.400	1.92	
REMI + Ph	0.356	1.27	
CP + Ph&BC (ours)	0.583	3.28	
CP + BC	0.461	2.25	
CP + Ph	0.306	1.78	
POP909 _{cut}		1.29	
Random Answer	0.130	2.50	

Add Phrase & Bar Countdown

- Our methods are effective in controlling phrase lengths
- Both Phrase and Bar Countdown events are necessary

Conclusion

Phrase lengths control with a sequential generation policy by

- Adding Phrase and BAR COUNTDOWN events to REMI and
- Extending autoregressive generation to **reflect** user input