

Abstract

Goal

To control phrase lengths with a sequential generation policy

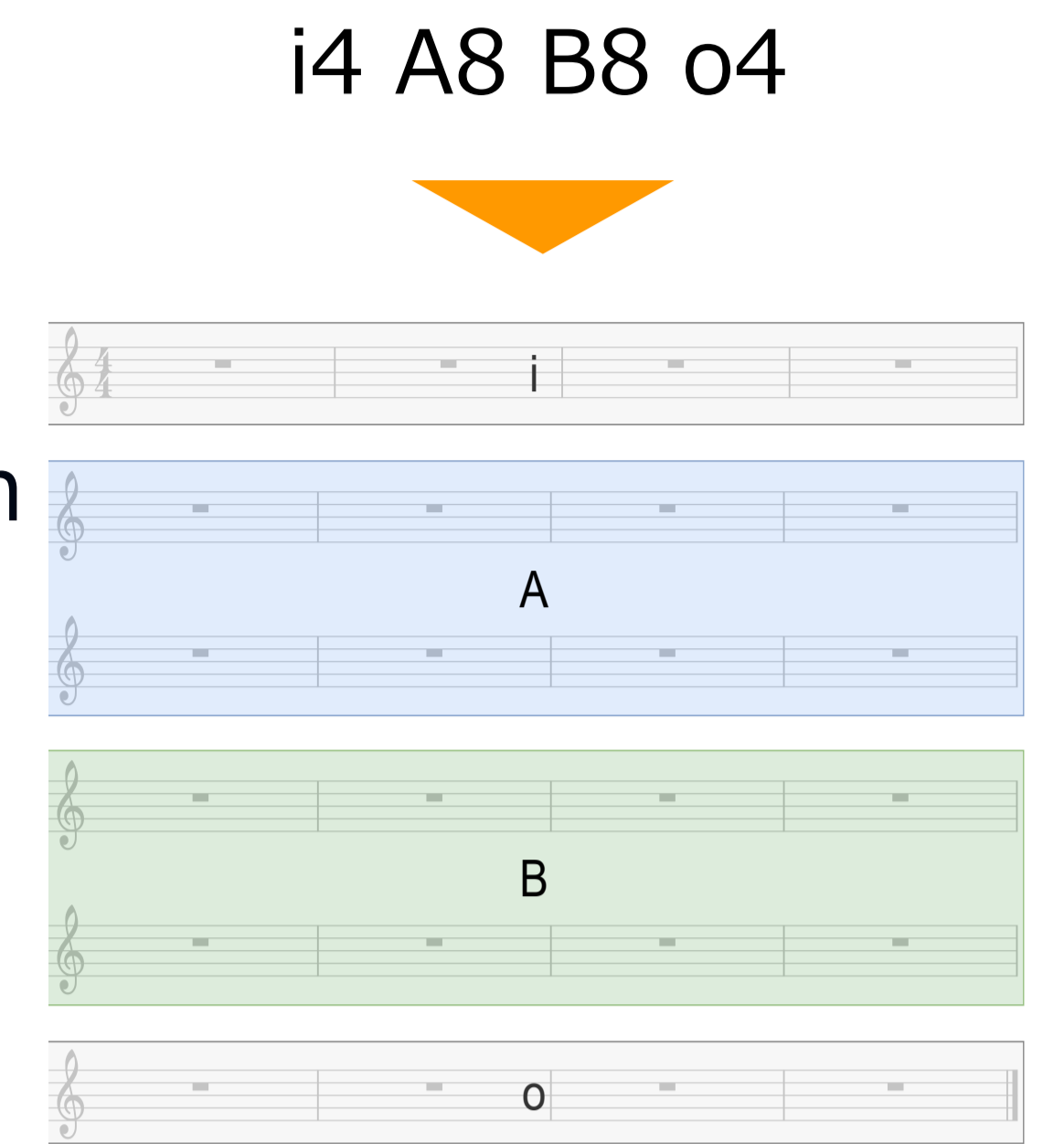
Input: Phrase configuration (label and length of each phrase)

Output: A piece with **designated phrase lengths** {
• **Phrase switching** at designated position
• **Natural closure** at the end of outro

Contribution

- We added **PHRASE and BAR COUNTDOWN events** to the existing event-based music representations (REMI [1] and CP [2]) and showed that **both events are necessary**
- We extended autoregressive generation to **reflect user input**

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



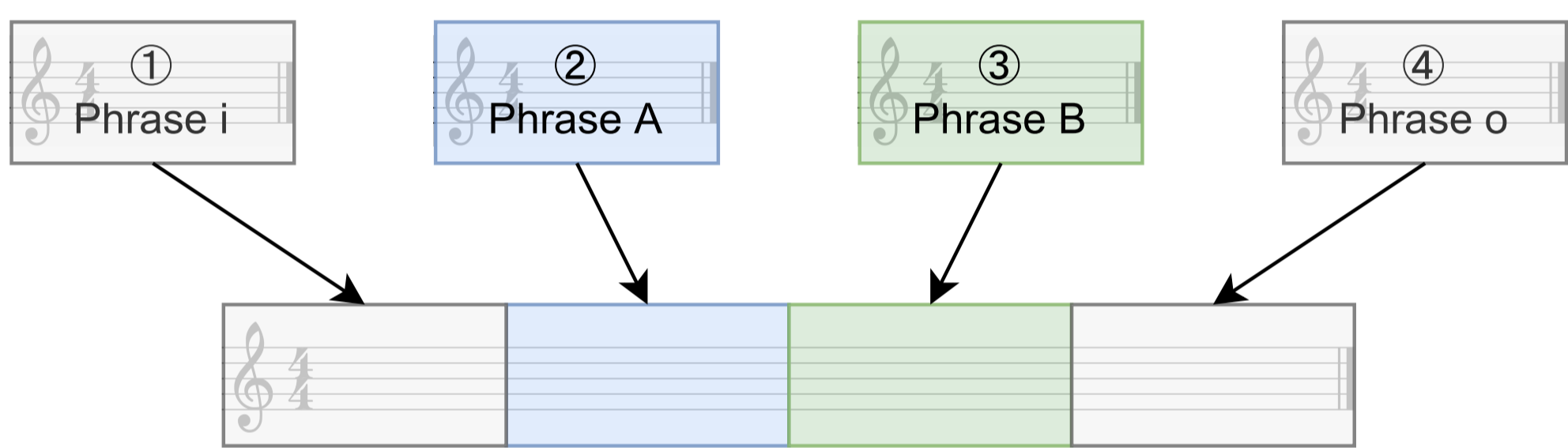
Introduction

Motivation

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

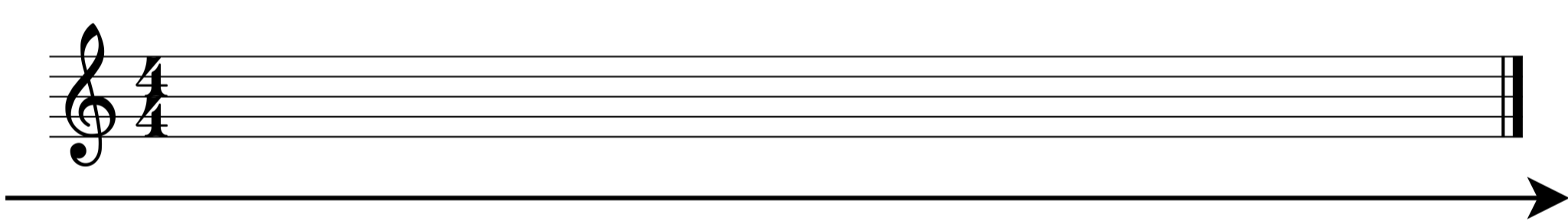
Two Strategies

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

Method

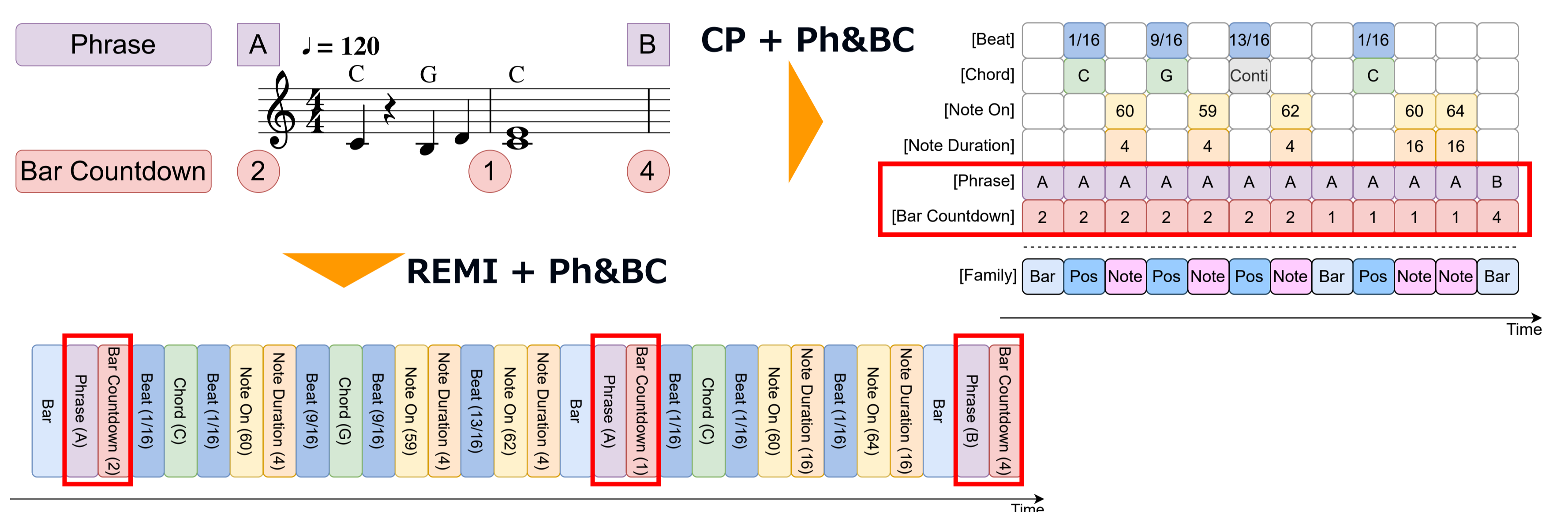
PHRASE and BAR COUNTDOWN Events



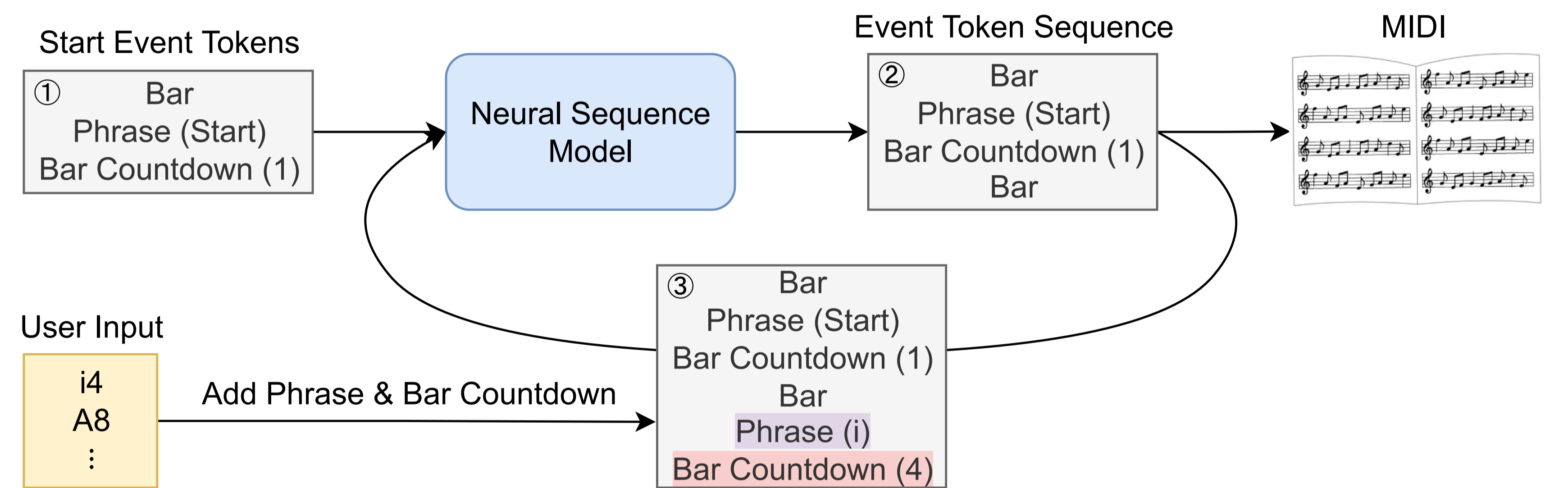
PHRASE (Ph): Which phrase a bar belongs to

BAR COUNTDOWN (BC): The number of bars remaining in a phrase

Examples of Event-Token Sequence



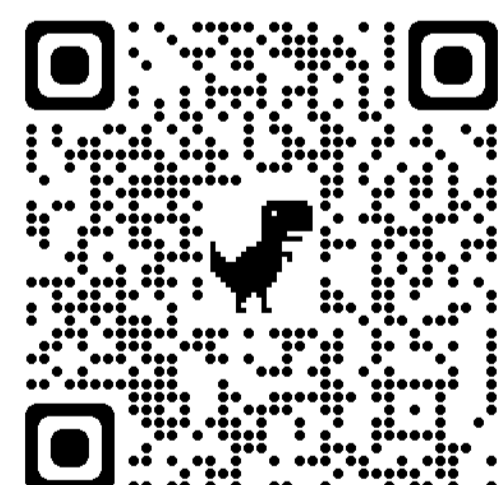
Generation Process — How to Reflect User Input?



- Start event tokens are input to the model to start generation
- Next event tokens are predicted by the model
- PHRASE and BAR COUNTDOWN are added based on user input

Generated Pieces

Samples of the generated pieces are available at



Experiments

Listening Tests

Test 1: Phrase switching at designated position

- Subjects identified 3 phrase boundaries

Test 2: Natural closure at end of outro

- Subjects answered whether closure was natural or abrupt

Results

- Our methods are **effective in controlling phrase lengths**
- Both PHRASE and BAR COUNTDOWN events** are necessary

Method	Test 1	Test 2
	Phrase Switching	Natural Closure
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

Conclusion

Phrase lengths control with a sequential generation policy by

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