Modeling The Rhythm From Lyrics For Melody Generation Of Pop Song

Daiyu Zhang*, Ju-Chiang Wang, Katerina Kosta, Jordan Smith, and Shicen Zhou
Speech, Audio, and Music Intelligence (SAMI) team

* daiyu.zhang@bytedance.com

SAMI | ISMIR 2022 BENGALURU | ByteDance 字节跳动

Background

- Creating a pop song melody according to pre-written lyricsis is a typical practice for composers.
- Lyrics and melody are not arbitrarily combined.
  - Prior analysis [1] indicates that patterns in lyrics and melodies are related and can be modeled, in part, with features of the melody (e.g., note duration) and lyrics (e.g., syllable stress).
  - An earlier lyric-to rhythm system also leveraged a dataset of pre-composed rhythms that were scored based on their match to the input syllable-stress and word-rarity patterns [2]
- Conditioning the choice of parameters in a generative model on data from other modalities, such as a bassline or a structure, can yield controllable generation systems [3]

System Overview
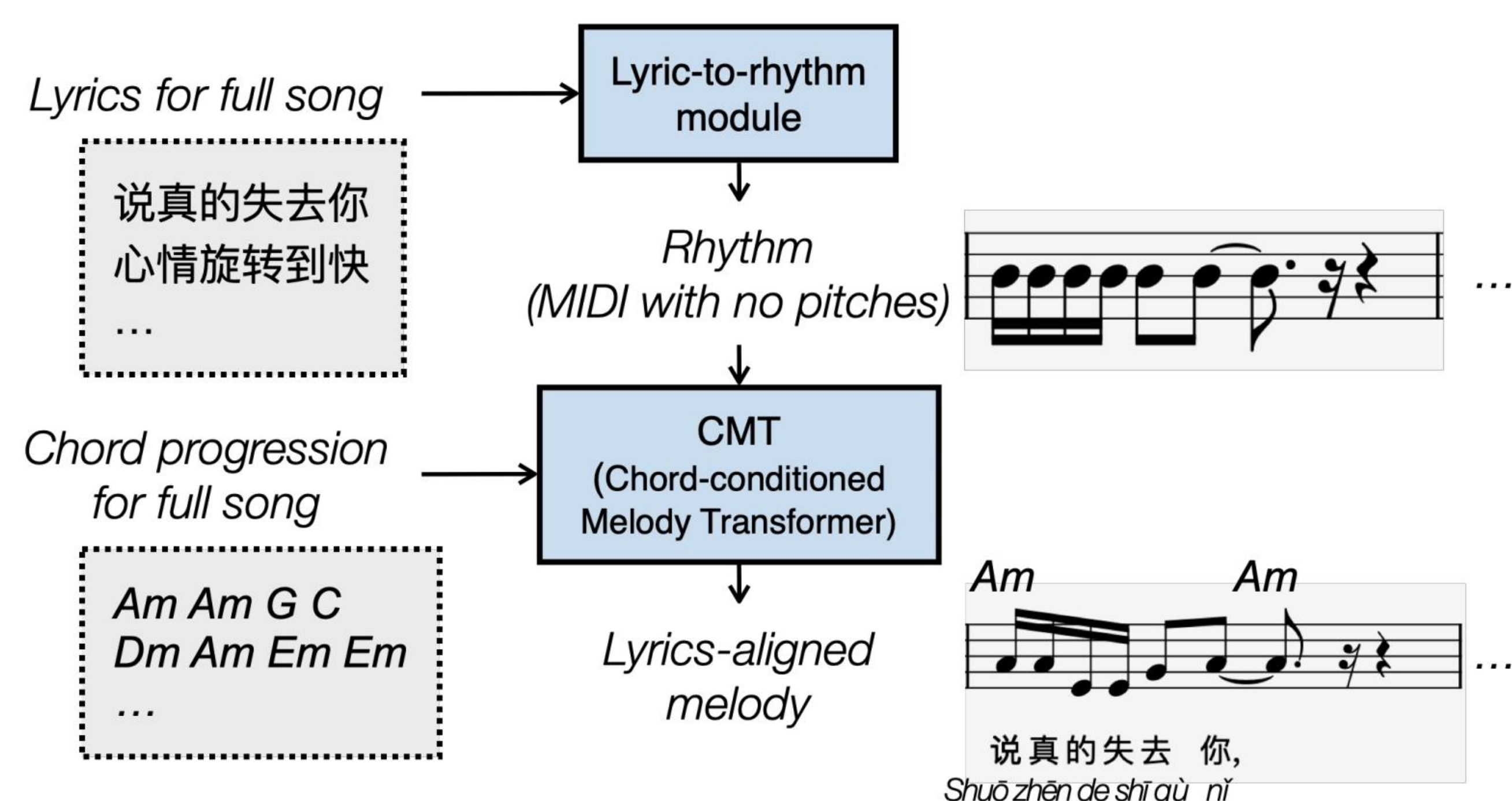
Reference

[1] E. Nichols, D. Morris, S. Basu, and C. Raphael, "Relationships between lyrics and melody in popular music," in *Proc. ISMIR*, 2009, pp. 471-476.
[2] E. Nichols, "Lyric-based rhythm suggestion," in *Proc. International Computer Music Conference*, 2009.
[3] J.-P. Briot, G. Hadjeres, and F. D. Pachet, *Deep learning techniques for music generation*. Springer Cham, 2020.
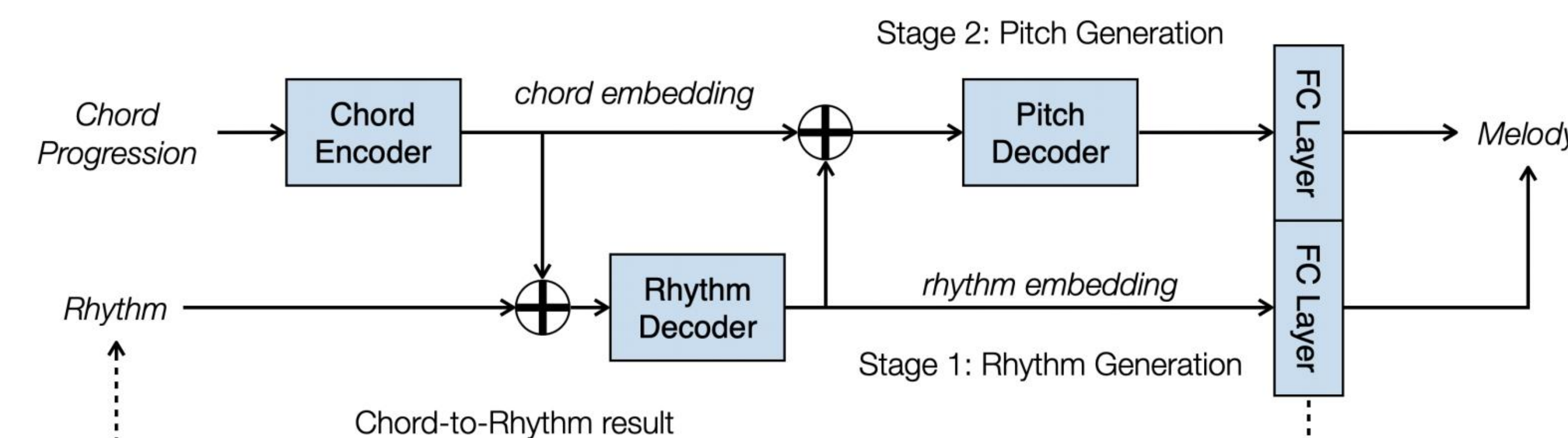
lyric-to-rhythm model
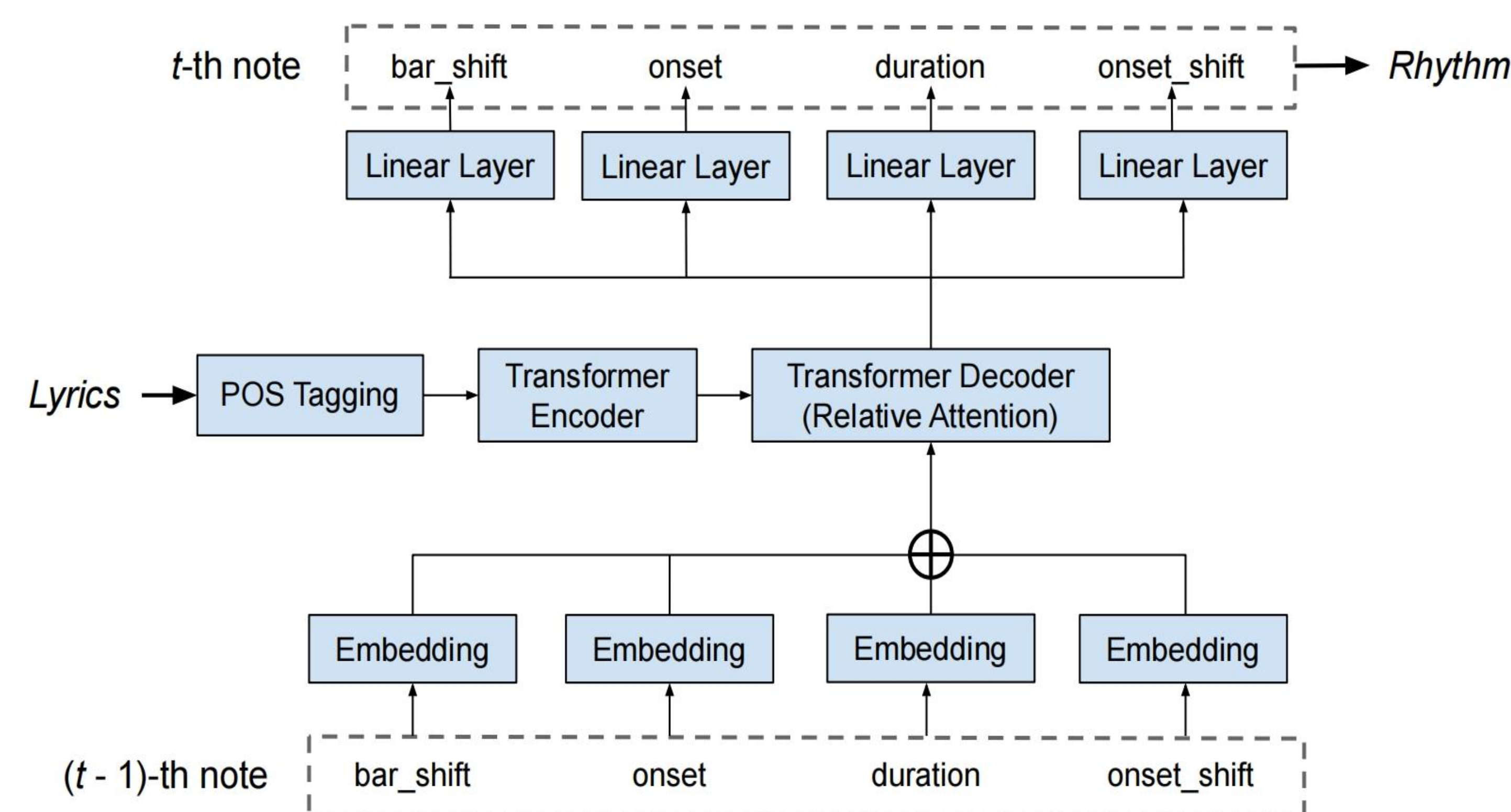
POS Tagging
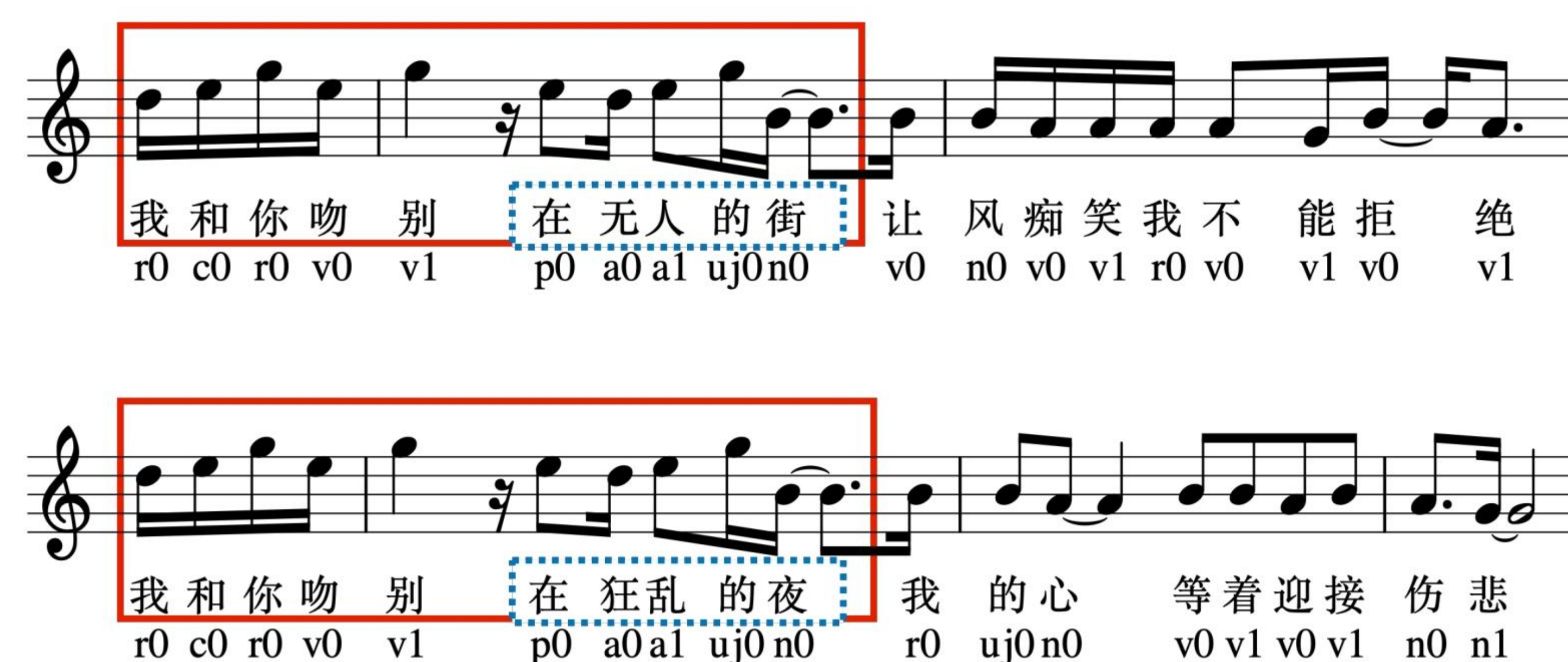
Rhythmic features for a note

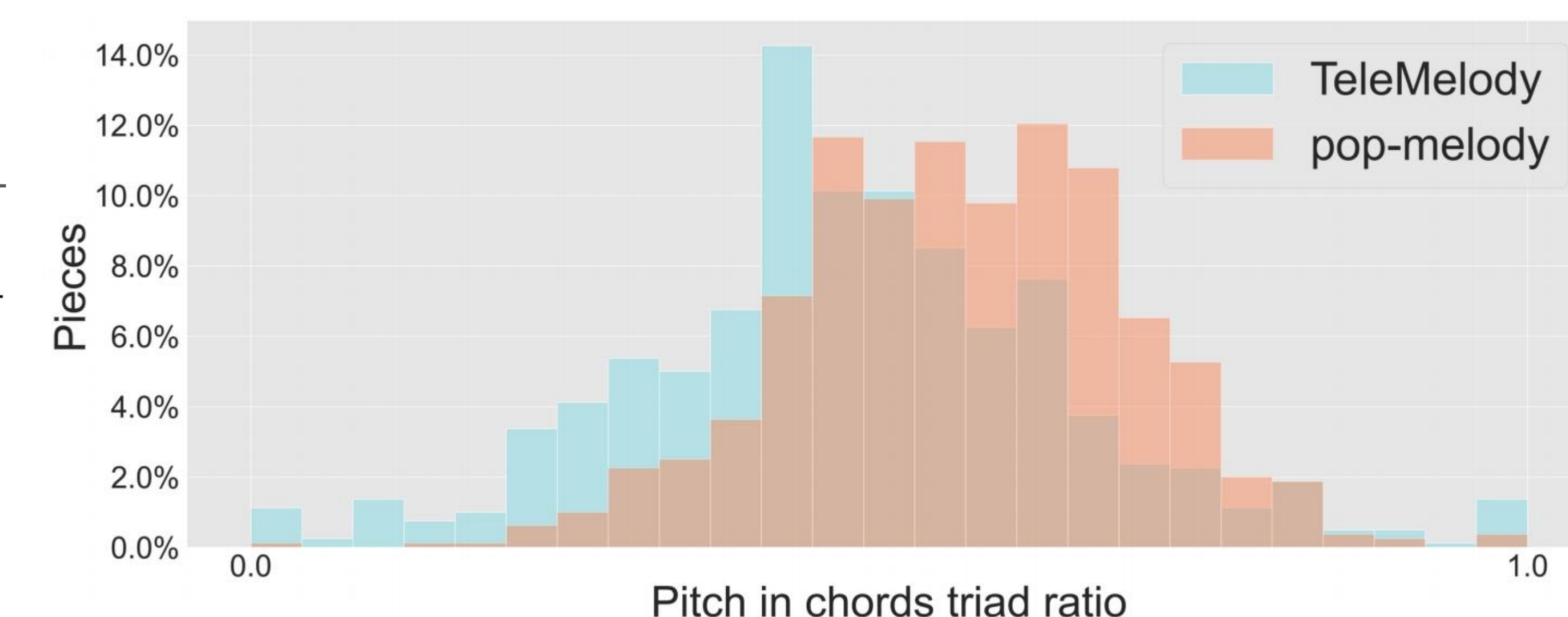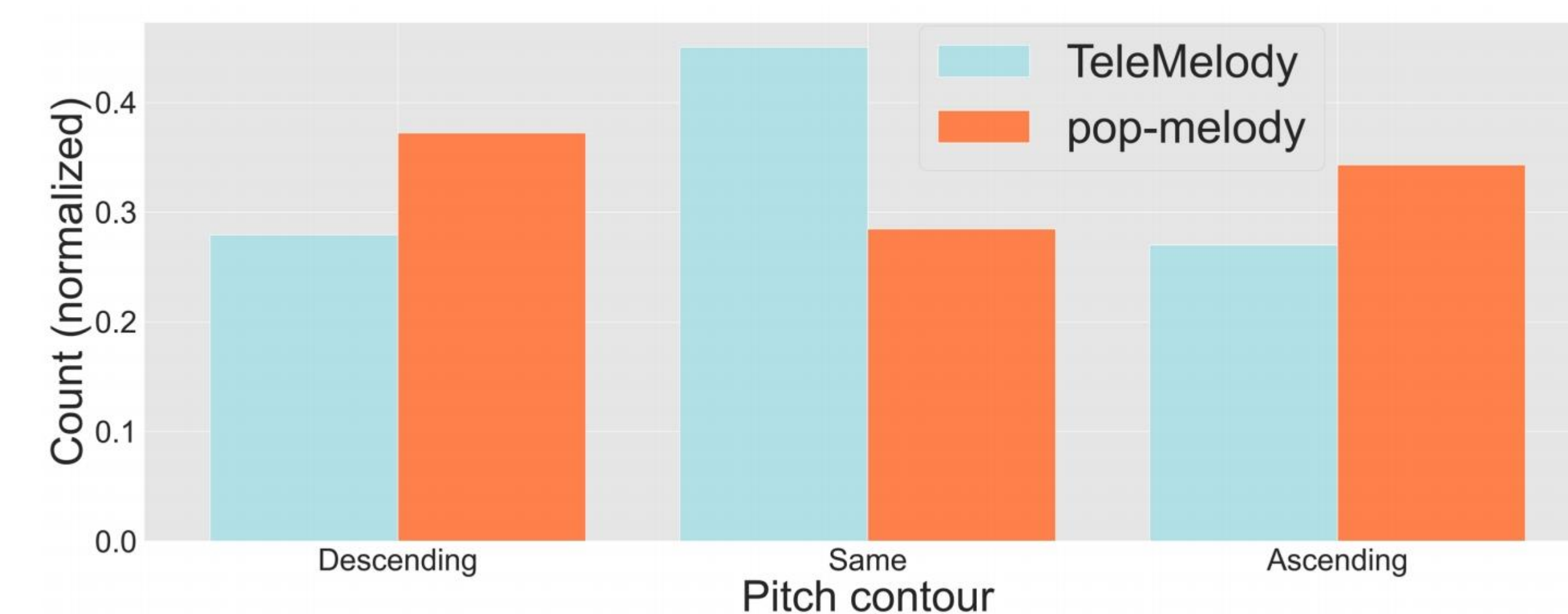| Token Name	Vocab.	Description
bar_shift	0, 1, 2	Time shift in bar to current bar
onset	0 – 15	Onset in 1/4 beat in current bar
duration	0 – 31	Duration in 1/4 beat
onset_shift	0 – 15	Time shift in 1/4 beat to previous note's onset

A two-stage structure of CMT

Subjective evaluation

System	Rhythm	Harmony	Melody	Overall
Baseline	3.42(.93)	3.67(.86)	3.46(.93)	3.42(.82)
TeleMelody	3.58(.96)	3.69(.90)	3.38(.85)	3.57(.73)
Pop-melody	3.84(.87)	3.87(.81)	3.64(.89)	3.68(.72)

Objective evaluation