

National

Taiwa

AUTOMATIC COMPOSITION OF GUITAR TABS BY TRANSFORMERS AND GROOVE MODELING

陳宥華 (Yu-Hua Chen)^{1,2,3}, 蕭文逸(Wen-Yi Hsiao)¹, 謝祖匡(Tsu-Kuang Hsieh)¹ 張智星(Jyh-Shing Roger Jang)³,楊奕軒 (Yi-Hsuan Yang)^{1,2} ¹Taiwan Al Labs, Taipei, Taiwan ²Academia Sinica, Taipei, Taiwan ³National Taiwan University



Motivation

- Automatic music composition
 - Describe piano music as a sequence of event tokens.
 - Representation for tabulature data are not yet be explored.
- Grooving
 - The best way to represent higher-level information for automatic composition is also unclear, especially for implicit information such as grooving.



Data acquisition

- 1. Compile our own guitar tab dataset with specific genre of fingerstyle
- 2. Data filtering
 - a. non-standard tuning
 - b. more than one guitar
 - c. low quality (wrong fingering and obvious annotation errors)



Backbone model

- Transformer-XL
 - Recurrence mechanism enable transformer model to capture relative mechanism for long-term dependency between each token.
 - Shows better result in previous music generation paper [1].

[1] Y.-S. Huang and Y.-H. Yang. Pop Music Transformer: Beat-based modeling and generation of expressive Pop piano compositions. In Proc. ACM International Conference on Multimedia, 2020.



Event representation





Grooving

- 1. Hard grooving
- 2. Soft grooving
- 3. Multi-resolution grooving





Evaluating the models

- Objective evaluation
 - \circ On fingering
 - \circ On grooving
- Subjective evaluation (user study)
 - \circ On comparison with real tabs



Objective evaluation (fingering)

	string (high-pitched \leftrightarrow low-pitched)							
	1st	2nd	3rd	4th	5th	6th		
(a) accuracy	100%	99%	97%	94%	91%	90%		
(b) pitch 42	$\sim 0\%$	$\sim 0\%$	10%	$\sim 0\%$	27%	63%		
(c) pitch 57	$\sim 0\%$	6%	65%	26%	$\sim 0\%$	$\sim 0\%$		
(d) pitch 69	85%	14%	$\sim 0\%$	$\sim \! 0\%$	$\sim 0\%$	$\sim 0\%$		



Objective evaluation (grooving)

	Hard ac	curacy †	Soft distance \downarrow		
	mean	max	mean	min	
hard grooving	76.2%	82.4%	56.3	44.6	
soft grooving	76.9%	83.0%	56.2	43.7	
multi-hard	79.0%	85.7%	57.8	44.3	
multi-soft	74.6%	81.1%	64.7	52.9	
no grooving	70.0%	80.1%	58.6	47.7	
training data	82.1%	89.5%	43.8	28.6	
random	64.9%	71.3%	70.6	59.6	



Subjective evaluation





Conclusions and future work

- New representation for tabulature data.
- Series of evaluations supporting the effectiveness of a modern neural sequence mode for higher level music information integration.



Audio samples and video

