AUTOMATIC COMPOSITION OF GUITAR TABS BY TRANSFORMERS AND GROOVE MODELING

Yu-Hua Chen[1,2,3], Yu-Siang Huang[1], Wen-Yi Hsiao[1] and Yi-Hsuan Yang[1,2]
[1]Taiwan AI Labs, Taiwan [2]Academia Sinica, Taiwan [3]National Taiwan University

ISMIR MTL2020

Motivation

Automatic music composition

- Describe piano music as a sequence of event tokens.
- Representation for tabulature data are not yet explored.

Grooving

- The best way to represent higher-level information for automatic composition is also unclear, especially for implicit information such as grooving.

Dataset

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

Backbone Model

We use **Transformer-XL** which shows better result in previous music generation paper.

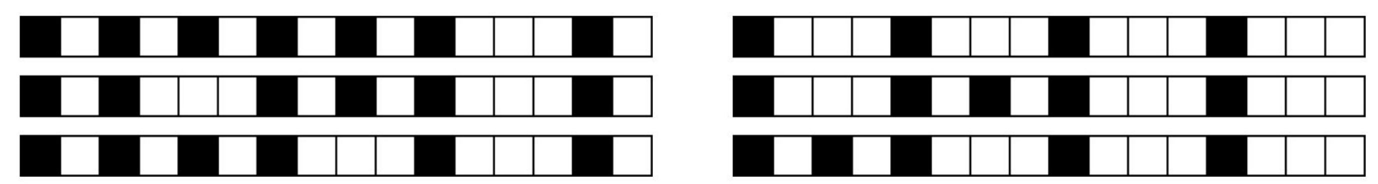
Audio samples and video

We provide additional audio samples and the video recording that is a guitarist from our team playing a generated tab in the QR code.

Grooving

Grooving of each bar are represented as a 16-dim vector.

- Hard grooving
- Soft grooving
- Multi-resolution grooving.

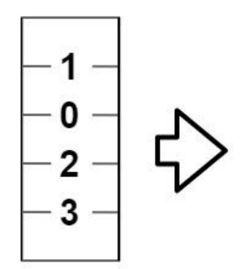


(a) (b)

Figure 3. Samples of 16-dim hard grooving patterns assigned to 2 different clusters (a), (b) by $k$means clustering.

Event Representation

Tab



Event Sequence

Bar,
Position_1/16, Note Velocity(16), Note On(48), Note Duration (8), String(5), Fret(3),
Position_1/16, Note Velocity(14), Note On(52), Note Duration (6), String(4), Fret(2),
Position_1/16, Note Velocity(16), Note On(55), Note Duration (4), String(3), Fret(0),
Position_1/16, Note Velocity(14), Note On(60), Note Duration (4), String(2), Fret(1),
...

Subjective Evaluation

We conduct a user study to ask the user to rate the generate tabs and real tabs.

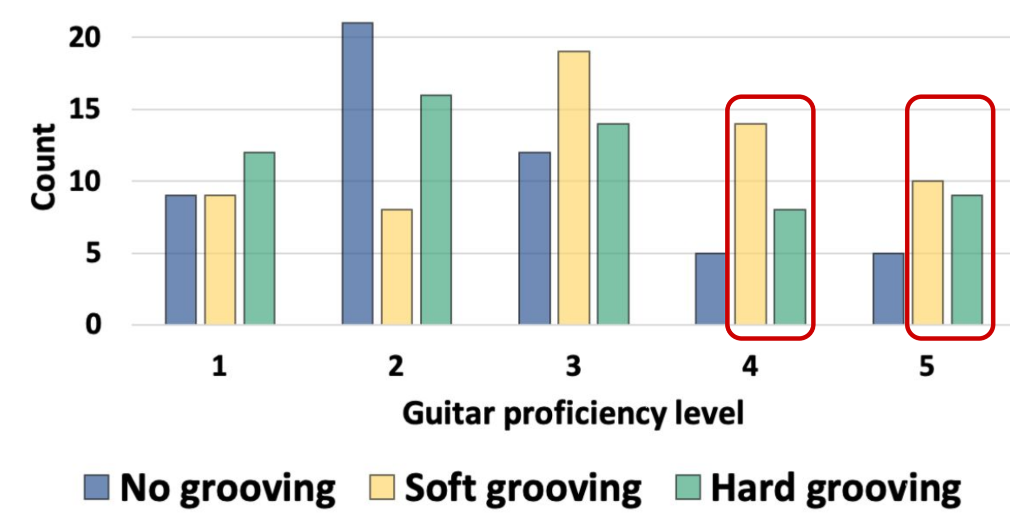


Figure 6. Result of the first user study asking subjects to choose the best among the three continuations generated by different models, with or without GROOVING, given a man-made prompt. The result is broken down according to the self-report guitar proficiency level of the subjects.

Objective Evaluation

On Fingering

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

Table 3. (a) The average accuracy of our model in associating each STRING with a NOTE-ON, broken down by string; (b–d) The string-relevant output probability estimated by our model for three different pitches.

On Grooving

We compare the performance of models trained with or without GROOVING for generating "continuations" of a given "prompt."

	Hard accuracy ↑		**Soft distance ↓**	
	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

Table 4. Objective evaluation on groove coherence.

Key Contributions

1. Proposed a new representation for tabulature data and grooving.
2. Proposed several evaluation methods for tab generation and grooving consistency.

AUTOMATIC COMPOSITION OF GUITAR TABS BY TRANSFORMERS AND GROOVE MODELING

Yu-Hua Chen[1,2,3], Yu-Siang Huang[1], Wen-Yi Hsiao[1] and Yi-Hsuan Yang[1,2]
[1]Taiwan AI Labs, Taiwan [2]Academia Sinica, Taiwan [3]National Taiwan University

ISMIR MTL2020

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.

Dataset

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

Audio samples and video

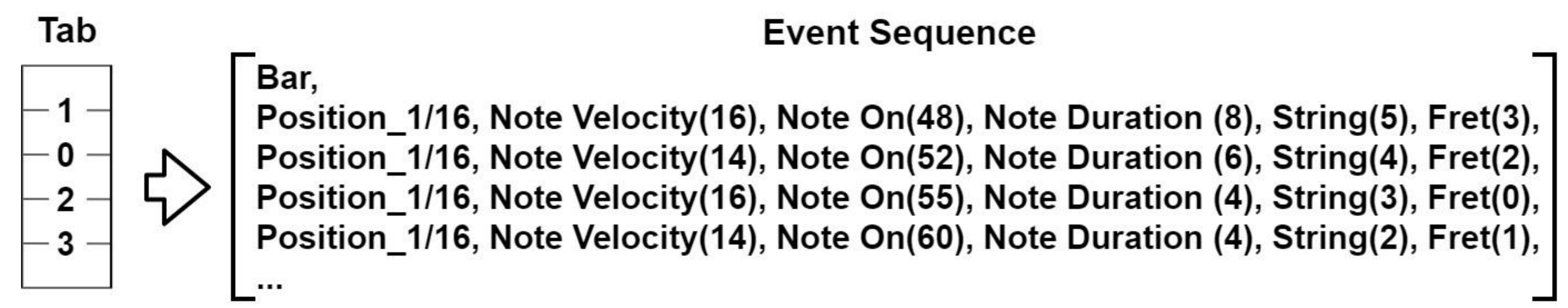
We provide additional audio samples and the video recording that is a guitarist from our team playing a generated tab in the QR code.

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.

Event Representation



Grooving

Grooving of each bar are represented as a 16 grids vector.

- Hard grooving
- Soft grooving
- Multi-resolution grooving.

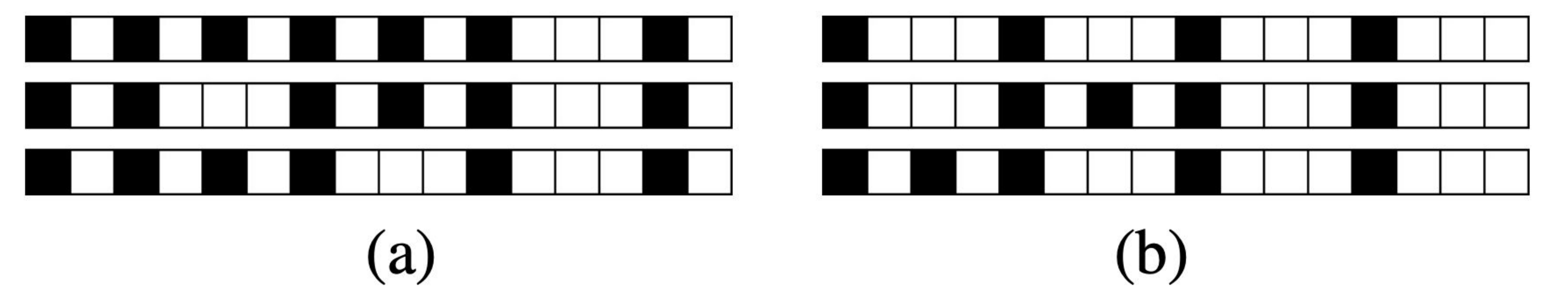


Figure 3. Samples of 16-dim hard grooving patterns assigned to 2 different clusters (a), (b) by *k*means clustering.

Experiment

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

Table 3. (a) The average accuracy of our model in associating each STRING with a NOTE-ON, broken down by string; (b–d) The string-relevant output probability estimated by our model for three different pitches.

	Hard accuracy ↑		Soft distance ↓	
	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

Table 4. Objective evaluation on groove coherence.

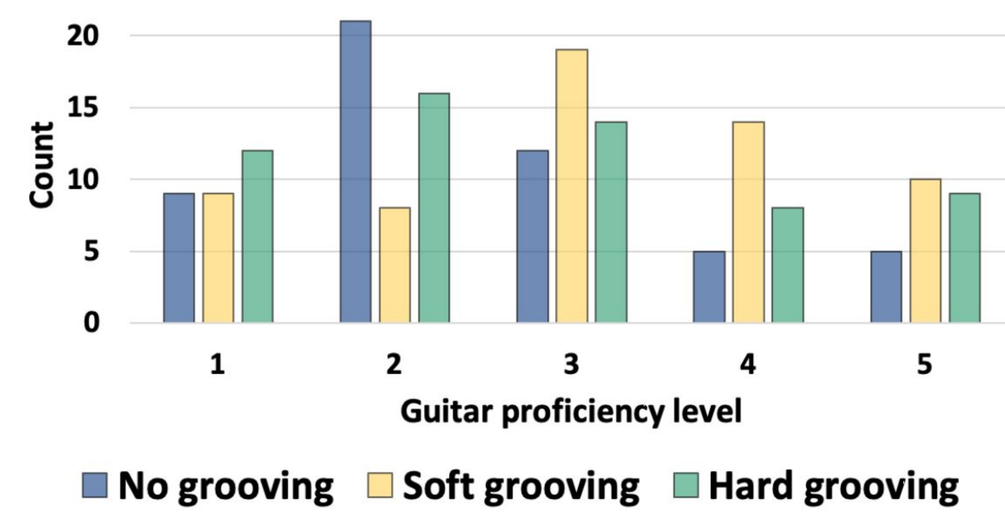


Figure 6. Result of the first user study asking subjects to choose the best among the three continuations generated by different models, with or without GROOVING, given a man-made prompt. The result is broken down according to the self-report guitar proficiency level of the subjects.

Key Insights

1. We proposed a new representation for tabulature data and grooving..
2. We provide series of evaluations supporting the effectiveness of a modern neural sequence mode for higher level music information integration.

AUTOMATIC COMPOSITION OF GUITAR TABS BY TRANSFORMERS AND GROOVE MODELING

Yu-Hua Chen[1,2,3], Yu-Siang Huang[1], Wen-Yi Hsiao[1] and Yi-Hsuan Yang[1,2]
[1]Taiwan AI Labs, Taiwan [2]Academia Sinica, Taiwan [3]National Taiwan University

ISMIR MTL2020

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.

Dataset

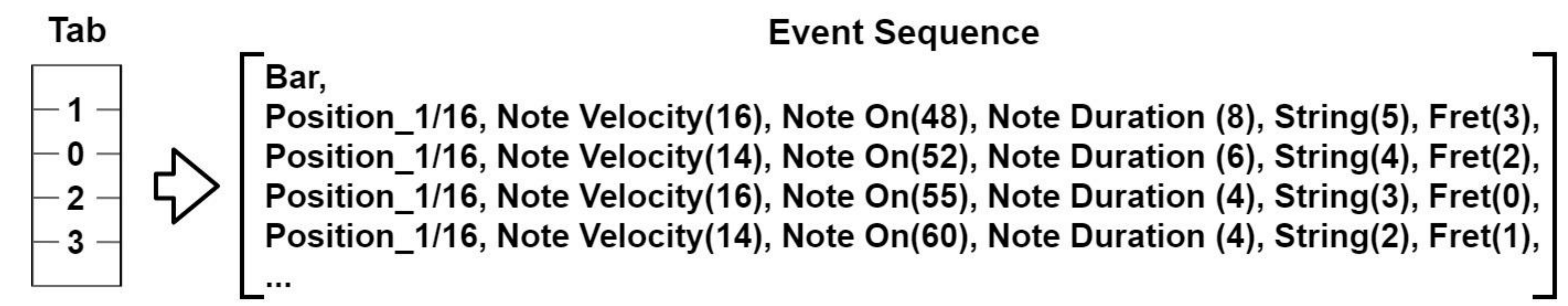
- Compile our own guitar tab dataset with specific genre of fingerstyle
- Data filtering:
  a. non-standard tuning
  b. more than one guitar
  c. low quality (wrong fingering and obvious annotation errors)

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].

Event Representation



Grooving

Grooving of each bar are represented as a 16 grids vector.

- Hard grooving
- Soft grooving
- Multi-resolution grooving.

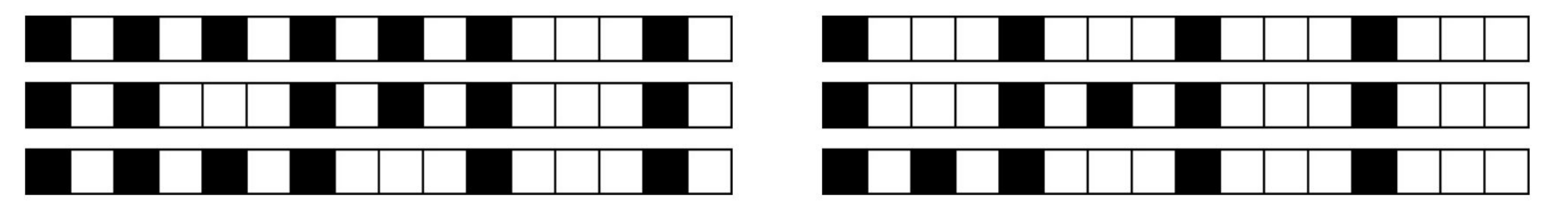


Figure 3. Samples of 16-dim hard grooving patterns assigned to 2 different clusters (a), (b) by *k*means clustering.

Experiment

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

Table 3. (a) The average accuracy of our model in associating each STRING with a NOTE-ON, broken down by string; (b–d) The string-relevant output probability estimated by our model for three different pitches.

	Hard accuracy ↑		Soft distance ↓	
	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

Table 4. Objective evaluation on groove coherence.

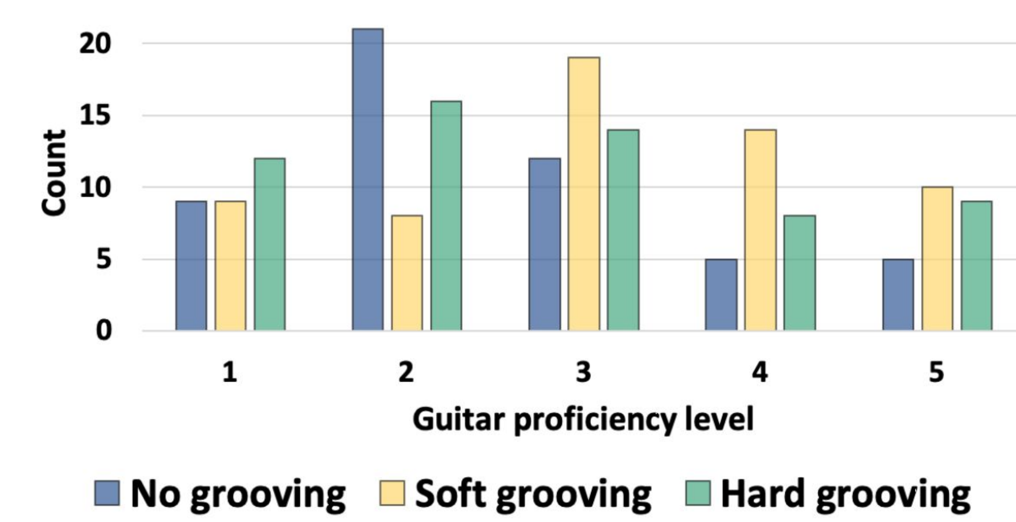


Figure 6. Result of the first user study asking subjects to choose the best among the three continuations generated by different models, with or without GROOVING, given a man-made prompt. The result is broken down according to the self-report guitar proficiency level of the subjects.

Reference

[1] Jesse Engel, Kumar Krishna Agrawal, Shuo Chen, Ishaan Gulrajani, Chris Donahue, and Adam Roberts. GANSynth: Adversarial neural audio synthesis. In Proc. ICLR, 2019.
[2] Bryan Wang and Yi-Hsuan Yang. PerformanceNet: Score-to-audio music generation with multi-band convolutional residual network. In Proc. AAAI, 2019

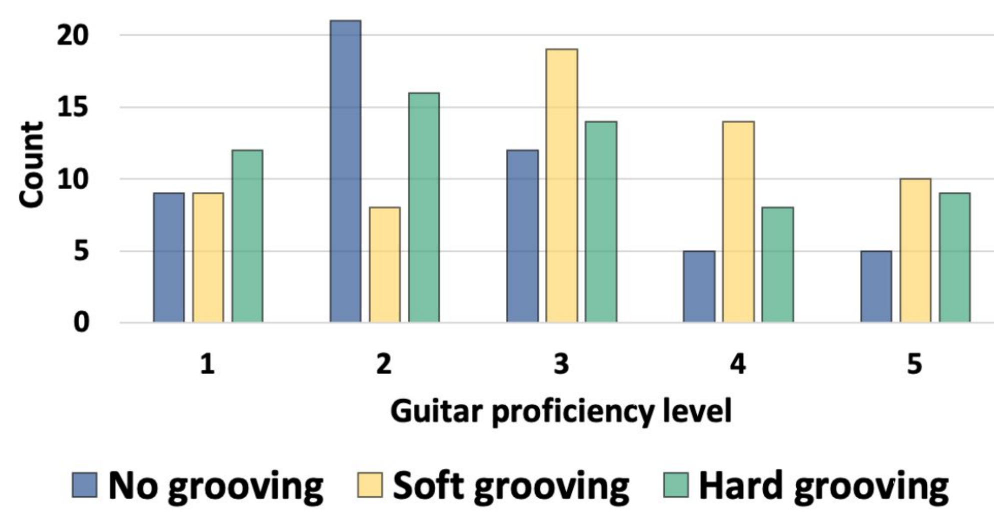


Figure 6. Result of the first user study asking subjects to choose the best among the three continuations generated by different models, with or without GROOVING, given a man-made prompt. The result is broken down according to the self-report guitar proficiency level of the subjects.

Objective Evaluatin

Key Insights

1. We proposed a new representation for tabulature data and grooving.
2. We provide series of evaluations supporting the effectiveness of a modern neural sequence mode for higher level music information integration.