DMRN+18: DIGITAL MUSIC RESEARCH NETWORK
ONE-DAY WORKSHOP 2023

QUEEN MARY UNIVERSITY OF LONDON
TUE 19 DECEMBER 2023

Automatic Guitar Transcription with a Composable Audio-to-MIDI-to-Tablature Architecture

Xavier Riley*, Drew Edwards and Simon Dixon

Centre for Digital Music, Queen Mary University of London, United Kingdom, j.x.riley@qmul.ac.uk

***Abstract*— This work-in-progress demonstrates an end-to-end guitar transcription system. The architecture takes as input a solo guitar recording, transcribes the audio to MIDI, and then estimates a tablature for the performance. The audio-to-MIDI transcription exhibits strong generalisability, including state-of-the-art performance on GuitarSet in a zero-shot setting. The tablature estimation is a novel approach applying masked language modeling to per-note string assignment.**

***Index Terms*—** guitar, transcription, tablature, AMT

I. GUITAR MULTI-PITCH ESTIMATION

Automatic transcription of piano has achieved good results in recent years due to the availability of large datasets such as MAESTRO [1]. Several successful architectures have been proposed, however the guitar does not yet have a comparable dataset with which to train these models. Existing guitar datasets tend to be smaller, with less timbral diversity [2]. We address this lack of data by adapting a recent score alignment technique proposed by Maman and Bermano [3]. We use this to produce aligned MIDI for 78 commercially available guitar recordings. These form our new dataset which we then use to fine-tune an existing piano model. In contrast the work by Maman and Bermano, we use a newer high-resolution piano model proposed by Kong et al. [4] which is shown to be more robust to noisy labels. We also use data augmentations on the MAESTRO dataset when training the base piano transcription model. This helps with generalisability when fine-tuned on guitar recordings.

II. TABLATURE ESTIMATION

Our approach to guitar tablature estimation uses the MIDI as input instead of audio. This loses timbral information but affords certain advantages. First, since the input and output are symbolic, a user can change the string and fret assignment of a particular set of notes and regenerate the estimated tablature. Second, this modular architecture provides a novel solution to arranging for guitar with a MIDI keyboard. A composer or arranger can play MIDI and quickly view how it could be performed on guitar.

We model the task of guitar tablature estimation as a masked language modeling task. Our ground truth data consists of guitar tablature transcriptions (from the 78 performances mentioned in Section I and GuitarSet [5]), in MusicXML or GuitarPro format. These are converted to six-track MIDI files, with one track per string. We use the Structured tokenizer from MidiTok [6]. For each note event N_i, we output the following tokens: $N_i \rightarrow S_i, T_i, P_i, V_i, D_i$, where $S_i \in \{1, 2, 3, 4, 5, 6\}$ is the string, T_i is the relative time shift, P_i is the pitch, V_i is the velocity, and D_i is the duration. During training, we mask and predict the S_i tokens. Our current best model using a BART [7] Transformer architecture achieves approximately 80% accuracy on a held out test set without any post-processing and is still a work-in-progress.

III. REFERENCES

[1] C. Hawthorne, A. Stasyuk, A. Roberts, I. Simon, C. A. Huang, S. Dieleman, E. Elsen, J. H. Engel, and D. Eck, "Enabling factorized piano music modeling and generation with the MAESTRO dataset," in *7th International Conference on Learning Representations*, New Orleans, USA, 2019.

[2] Y. Chen, W. Hsiao, T. Hsieh, J. R. Jang, and Y. Yang, "Towards automatic transcription of polyphonic electric guitar music: A new dataset and a multi-loss transformer model," in *IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2022, pp. 786–790.

[3] B. Maman and A. H. Bermano, "Unaligned supervision for automatic music transcription in the wild," in *International Conference on Machine Learning, ICML 2022, 17-23 July 2022, Baltimore, Maryland, USA*, ser. Proceedings of Machine Learning Research, vol. 162. PMLR, 2022, pp. 14 918–14 934.

[4] Q. Kong, B. Li, X. Song, Y. Wan, and Y. Wang, "High-resolution piano transcription with pedals by regressing onset and offset times," *IEEE ACM Transactions on Audio, Speech and Language Processing*, vol. 29, pp. 3707–3717, 2021.

[5] Q. Xi, R. M. Bittner, J. Pauwels, X. Ye, and J. P. Bello, "GuitarSet: A dataset for guitar transcription," in *Proceedings of the 19th International Society for Music Information Retrieval Conference*, Paris, France, 2018, pp. 453–460.

[6] N. Fradet, J.-P. Briot, F. Chhel, A. El Fallah Seghrouchni, and N. Gutowski, "MidiTok: A python package for MIDI file tokenization," in *Extended Abstracts for the Late-Breaking Demo Session of the 22nd International Society for Music Information Retrieval Conference*, 2021.

[7] M. Lewis, Y. Liu, N. Goyal, M. Ghazvininejad, A. rahman Mohamed, O. Levy, V. Stoyanov, and L. Zettlemoyer, "Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension," in *Annual Meeting of the Association for Computational Linguistics*, 2019.

*XR and DE are research students at the UKRI Centre for Doctoral Training in Artificial Intelligence and Music. XR is supported by UK Research and Innovation [grant number EP/S022694/1]; DE is supported by Queen Mary University of London and Yamaha.