DMRN+16: DIGITAL MUSIC RESEARCH NETWORK
ONE-DAY WORKSHOP 2021

QUEEN MARY UNIVERSITY OF LONDON
TUE 21 DECEMBER 2021

Beat-Based Audio-to-Score Transcription for Monophonic Instruments

Jingyan Xu

Music X Lab, NYU Shanghai, joy_xjy @ sjtu.edu.cn

***Abstract*— We propose a model to generate readable scores from audios for monophonic instruments in classical music. Firstly, we obtain the beats from the transcribed MIDI. Secondly, we analyze the most likely tone combinations and pitches according to the beats. Thirdly, we do the recreation to refine the potential mistakes in pitches and rhythms to make the musical semantics more reasonable. The generated scores are subjected to the performers' intentions and are meaningful in musicology.**

***Index Terms*—** audio-to-score, beat tracking, music generation

I. INTRODUCTION

Audio-to-score is to estimate the human-readable score from the input audio signal. For instruments with stable rhythms and fixed pitches like piano, there exist decent methods to obtain accurate scores [1]. However, for monophonic instruments, the unstable rhythms and the constantly changing pitches make the score difficult to obtain. As the performers add their own improvised recreations in real performances, the original scores and the recreated scores might be different. We want to stress this problem in our work.

We try to recover the performers' intentions into human-readable scores. To acquire the performance scores, people first estimate the tones, beats, rhythms, and pitches by repeatedly listening to the recording. The ambiguity makes it impossible to obtain the rhythms and pitches accurately. In face of this problem, people may add their own recreations to make the score be reasonable in musicology. Our model does a similar job by stretching MIDI notes based on the extracted beats. After that, the model performs recreation based on music semantics.

Our model can generate scores as long as the transcribed information meets the minimum requirement [2]. With the common information in the part scores, a further recreation could lead to a readable full score for a band or an orchestra in the future.

II. METHODOLOGY

We primarily consider 8-measure music segments in 4/4 time signature. The tones are transposed to C major or A minor , and there are no off-key notes.

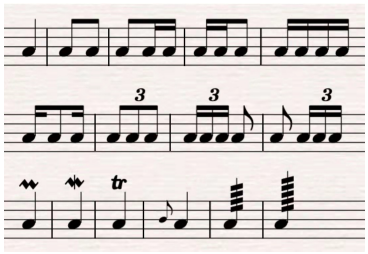

Figure 1: The possible combinations in one quarter note

Step 1: We extract beats from the MIDI, and MIDI are acquired from a transcription model.

Step 2: We jointly estimate the most possible tone value combinations, pitches, onsets and offsets.

We suppose that one beat is a quarter note. Therefore, the possible tone combinations for a quarter note are finite as Figure1 illustrates.

To distinguish between different note durations, we also need to label the onsets and offsets of the notes.

Step 3: We jointly refine the potential mistakes in pitches according to a music language model and fix the notes.

We use a public score editing software MuseScore 3 for score typesetting and generate the readable scores in the MusicXML format. For baselines, scores are generated by the MIDI data from Step 1 or Step 2, but not both. For evaluation, we use the mean of the 5 error rates in [3] to evaluate the quality of our generated scores. As there are some recreations in our results, the subjective evaluations are also indispensable.

III. REFERENCES

[1] Y. Hiramatsu, E. Nakamura, and K. Yoshii, "Joint estimation of note values and voices for audio-to-score piano transcription," in *Proceedings of the 22th International Society for Music Information Retrieval Conference (ISMIR)*, 2021.

[2] L. Lin, Q. Kong, J. Jiang, and G. Xia, "A unified model for zero-shot music source separation, transcription and synthesis," in *Proceedings of the 22th International Society for Music Information Retrieval Conference (ISMIR)*, 2021.

[3] E. Nakamura, E. Benetos, K. Yoshii, and S. Dixon, "Towards complete polyphonic music transcription: Integrating multi-pitch detection and rhythm quantization," in *2018 IEEE International Conference on*

Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2018, pp. 101–105.