Review on manuscript
Unsupervised Transcription of Piano Music

By Md Nur Amin

1 Summary

This paper outlines a model for transcribing piano music in an unsupervised fashion to address the source separation problem by learning recording-specific spectral profiles and temporal envelopes. Unlike existing models for symbolic music generation, this work addresses the problem of source separation efficiently by better modeling the discrete musical structure and adopting the timbral properties of piano at the same time.

Overall, I think the paper renders an interesting application where it efficiently couples the two stated approaches to achieve a notable result in unseen piano samples although a questionable performance on a varied piano sample.

2 Merits

1.Sound prior knowledge and comprehensive representation of the model.

2.An effective approach concatenating of two contemporary aspect in this particular domain.

3 Critiques

1.Lack of explanation of some terms, modeling during the Activation and Learning and Inference stage.

2.Probable flaws during sampling the data and taking account of the sampling instrument.

4 Analysis

1.The authors seems to have a sound prior knowledge that helps to render a model that solves the discreet musical structure and timbral properties of piano at the same time. At the same time, pictorial depictions of different stages of the modeling make the paper easy to understand.

2.The paper is mostly comprehensive while an explanation of several terms, for instance, spectral profile, spectrum of harmonics and temporal shapes are absent.

3.Although skipping the block-coordinate ascent updates while updating the envelope and spectral parameters F1 does not drop significantly.

4.clear explanation is not provided in the Activation Model about the shape of rise and decay of parallel musical notes event associating to Event Model as play and rest respectively, regardless of their duration and velocity with envelope parameters.

5.The reason for feeding external corpus of symbolic music data in the distribution of notes in the original musical note was not specified during the Learning and Inference stage.

6. There is a lack of discussion about using the superposition instead of approximating MAP of the parameters.

7. Of the two corpus, MAP contains the sample sound of a particular acoustic pianist. Expressiveness varies from pianist to pianist depending on playing style. This phenomenon impacts the distribution of velocity in the activation segment. For the play event, the velocity distribution is unified in the activation segment. There is a great possibility this model will suffer when MIDI corpus generated from a different pianist will be employed.

8. With respect to the provided audio samples, It is about a certain type of "Disklavier" acoustic piano. I wonder how this model will perform on other types.

9. The input audio is represented as a magnitude spectrum short-time Fourier transform with a 4096-frame window and a hop size of 512 frames. Why exactly was this particular setting chosen?

10. I found it difficult to follow the narrative of the transition from Activation Model to Component Spectrogram Model and it persists during updating parameters between these two events in the Event Model section.

11. In section 4, during Initialization and Learning, the average of the parameter values across several synthesized pianos was taken into the account. Which MIDI corpus was used between the acoustic piano and MAPS corpus from the IMSLP library?