



Background

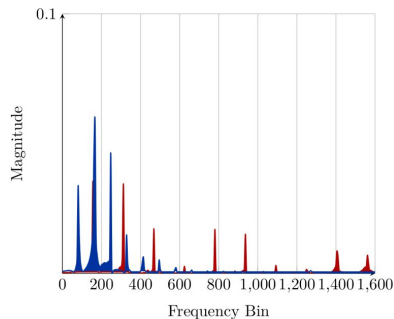
Automatic music transcription is a difficult task due to the complexity of musical audio. Creating an algorithm capable of profiling different instruments, recognizing individual notes in polyphonic music, and tracking beats or rhythms is a difficult problem of signal processing, with research on this topic spanning decades. We developed a solution using machine learning to automatically transcribe a given audio input, with limitations due to complexity.

Research

In academic literature, the problem of music transcription would be best solved through machine learning.^[1] This is done through sequence-to-sequence models, with latest advancements made using the transformer architecture.^[2] When given a waveform, there is currently no model that can transcribe sheet music with complete accuracy^[2]. However, there are multiple research papers that suggest a methodology that converts an audio source to MIDI^[1], which is then translated to MusicXML (a sheet music representation language)^[3]. Such a solution is primarily done by inputting data waveforms into a transformer model; by using deep learning, the model predicts an output sequence.

Implementation

Our implementation uses the JUCE framework (written in C++) in tandem with two machine learning models, MT3 and Score transformer (based on the transformer architecture) and Fast Fourier Transforms (FFT) to separate the waveforms. Such an implementation lets our application be modular and portable, as our goal is to create an easy-to-use experience. Using the JUCE framework allows for modularity in our implementation, as JUCE supports compilation as an independent application or a plugin. Such an approach allows our solution to be portable and compatible with beginner, enthusiast and professional workflows.



Electric Guitar playing notes E2 (blue) and E3 (red)

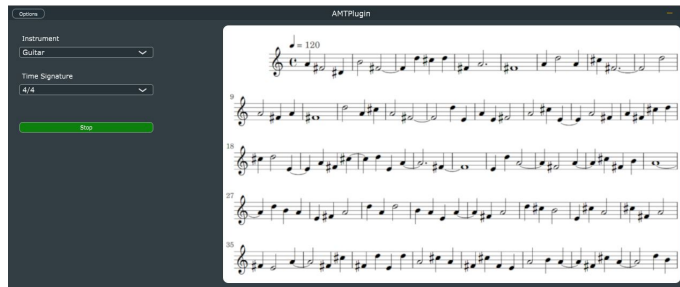
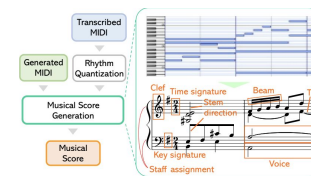
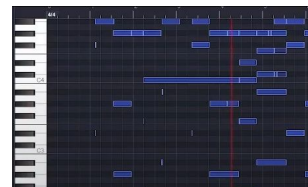
Each of an instrument's notes has a distinct frequency domain profile. Detecting each note and its duration can be done by checking for each note in small time slices across an audio input.



Initially, the MT3 (Multi-Task Multitrack Music Transcription) machine learning model is sent an audio input that is converted to MIDI.



Transcription to sheet music from MIDI is accomplished by passing the resulting MIDI file into the Score transformer model. This directly translates MIDI notes to sheet music representation language.^[1]



Our user interface is drawn through JUCE, which prompts the user to choose the instrument they want to transcribe and time signature. Once "Start" is clicked, the application transcribes the audio input and renders the sheet music.

Demographic

Our solution can be utilized by a wide spectrum of musicians, from beginner-level users that cannot write sheet music, enthusiasts transcribing for practice or later use, and professionals who want to increase their productivity. Music transcription takes time regardless of a person's musical literacy. Automatic music transcription would lower the barrier of entry to notating music.

Conclusion

Automatic music transcription is a challenging problem that is actively being researched. Due to these challenges, we had to restrict the audio input to a singular instrument. Manual transcription is tedious even for those with musical literacy and time. By using a combination of the FFT and current machine learning algorithms, we are able to create an application that can produce moderately accurate sheet music from a singular instrument.

[1] Suzuki, M., "Score Transformer: Generating Musical Score from Note-level Representation", <arXiv e-prints>, 2021. doi:10.48550/arXiv.2112.00355. <https://arxiv.org/abs/2112.00355>

[2] Yu-Siang Huang and Yi-Hsuan Yang. 2020. Pop music transformer: generating music with rhythm and harmony. In Proceedings of the 28th ACM International Conference on Multimedia

[3] Gardner, J., Simon, I., Manilow, E., Hawthorne, C., and Engel, J., "MT3: Multi-Task Multitrack Music Transcription", <arXiv e-prints>, 2021. doi:10.48550/arXiv.2111.03017.