Music Transcription System – A Review

Reddychakradhar Goud T.[1], Chandra Sekhar K.[2], Prashanth V.[3], Narasimha Deva G.[4], Deepika Ghai[5]

[1-4]*B. Tech Student, School of Electronics and Electrical Engineering, Lovely Professional University, Phagwara, Punjab, India – 144411*

[5]*Assistant Professor, School of Electronics and Electrical Engineering, Lovely Professional University, Phagwara, Punjab, India – 144411*

---***---

Abstract: This paper reviews the music transcription system and provides brief idea about music transcription along with various methods used to transcribe the music, those methods include Short Term Fourier Transform (STFT), Continuous Wavelet Transform (CWT) and Musical Instrument Digital Interface (MIDI) conversion. It also describes music information retrieval like feature extraction, pitch estimation, onset and offset detection, and frequency calculation. Current paper mainly focuses on reviewing monophonic music as it possess only one sound at a time, whereas polyphonic music is difficult to extract because of having multiple parameters involved at each time instance. Transcribing music is one of the challenging and emerging field in the digital signal processing, even in image processing.

Keywords: Short Term Fourier Transform, Continuous Wavelet Transform, Music Instrument Digital Interface, Digital Signal Processing, Image Processing.

1. INTRODUCTION

This paper reviews on developing an electronic device that generates an alphabetical music notes or music sheet that contains clefs from the music. Each key from a musical instrument produce unique frequency, this frequency is analysed and a particular alphabet (chromatic) from octave (G-A-B-C-D-E-F-G) is assigned to that particular key, which produced frequency. Each key in a music sample can be represented as symbols on a music sheet called as clefs. There are mainly 3 types of clefs. Namely treble clef, alto clef and bass clef. Most commonly seen clefs in a music sheet are treble and bass. On a piano, treble clef constitutes the keys that are played using right hand (basically keys that are present on right hand side), alto clef constitutes the keys that are played using both hands (basically keys that are present in the centre) and bass clef constitutes the keys that are played using left hand (basically keys that are present on left hand side). As we move from bass clef to treble clef frequency of the keys increases.
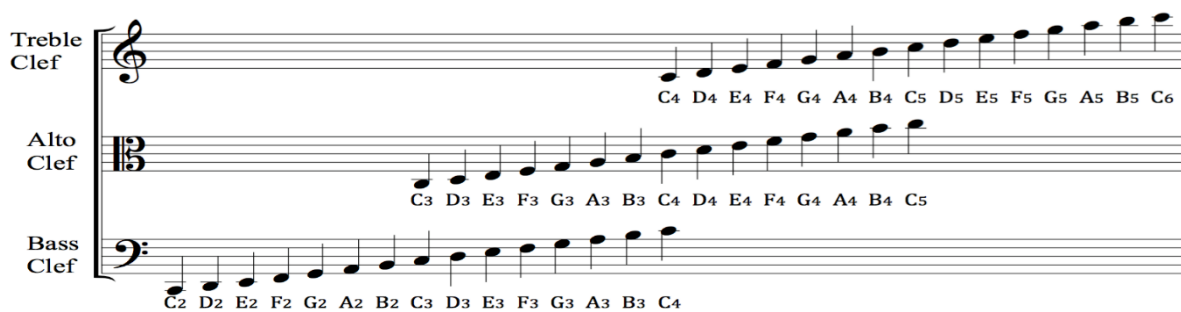

Fig. 1: Representation of clefs constituting keys on a piano

2. RELATED WORK

There are various methods with different techniques to transcribe music. To transcribe a music we have extract frequency, onsets, offsets, frequency and time. All these parameters can be extracted using Audio Content Analysis (ACA) [1]. There many others ways to transcribe music depending on the type of musical instrument and type of music like monophonic or polyphonic with different parameters [2]. Many techniques from digital signal processing like Short Term Fourier Transform, Discrete Wavelet Transform and Continuous Wavelet Transform can be used to evaluate the parameters like frequency, pitch, time duration as described in [3]. We can also use artificial neural networks using Long Short Term Memory (LSTM) for polyphonic music sequences for feature extracted as given in [4]. Besides these methods there other methods to transcribe music sheet using electronic device as described in [5]. For MIDI conversion there is a need of probability factors to determine the existence of musical sequence in a particular region [6].
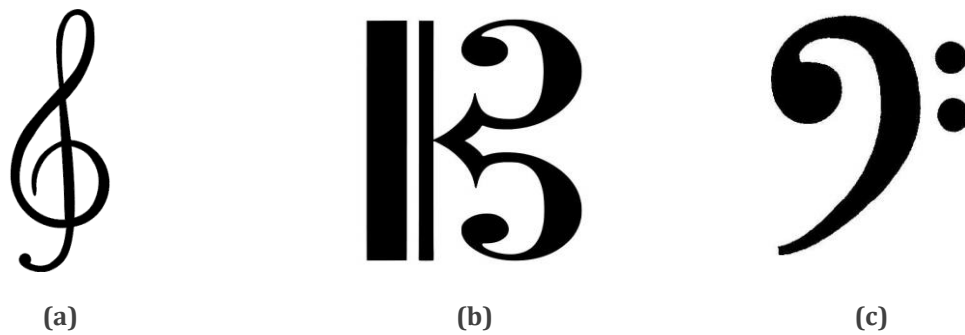
(a)   (b)   (c)

Fig. 2: Representation of clefs: (a) Treble clef; **(b)** Alto clef; **(c)** Bass clef

The current review on music transcription describe various methods in signal processing, those are helpful in evaluating various parameters required for fruitful music transcription. It also explains about various components like clefs, music sheet which are major components of music. Besides these it also reviews on MIDI conversion and ease in finding parameters from MIDI file.

Most of the methods are based on neural networks like using convolutional neural networks, recurrent neural networks, long short term memory and Q-transform for finding multiple pitches especially in polyphonic music as described in [7]. Similarly using deep neural networks in creating piano-roll as proposed in [8]. Few other ways are using graphical user interface (GUI) for feature extraction and transcription as in [9]. Support vector machine (SVM) can be used to determine the pitch as given in [10] and few methods for MIDI conversion like techniques given in [11]. All these techniques are essential in some way to extract required parameters from the music as described in [12].

3. PROPOSED METHODOLOGY

In this method we analyse the techniques music transcription using digital signal processing and MIDI conversion using various techniques.

3.1. EXTRACTION OF PARAMETERS

For extracting various parameters like frequency and time duration, we can use STFT and CWT for creating various graphs like scalogram, spectrogram and chromagram.
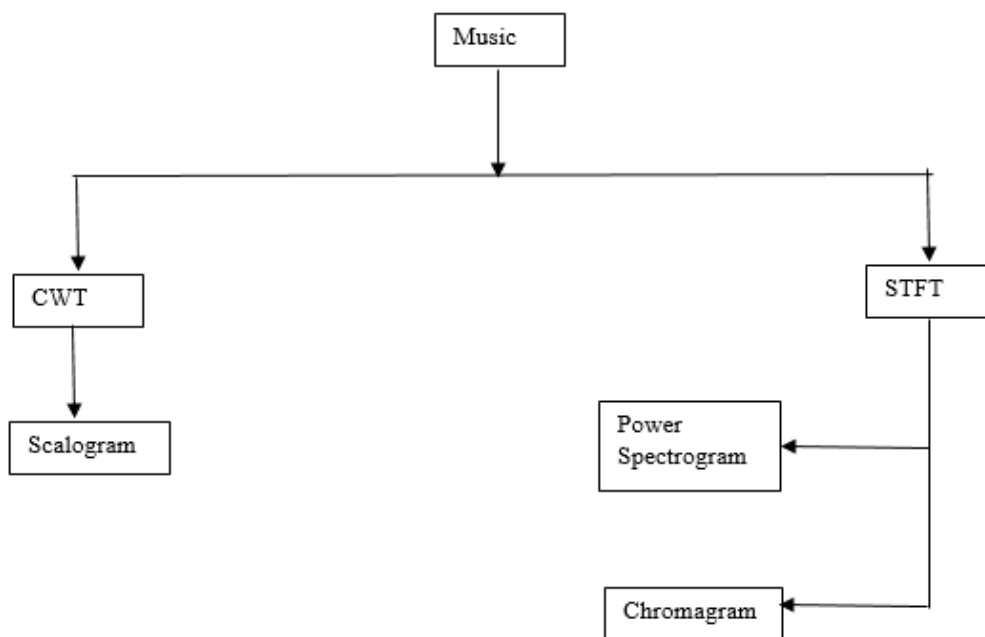

Fig. 3: Block Diagram for Parameter Extraction

When the music signal undergo CWT, it creates scalogram and it is analysed to evaluate frequency with respect to the time. Scalogram is the representation of a signal in a scale-time domain. CWT uses wavelets to get maximum resolution. Whereas STFT gives power spectrogram and chromagram. Power spectrogram is a frequency-time graph use to analyse amplitude of the signal. Finally by using chromagram, we can get note class of keys present in the music as the chromagram is the graph between note-class and time.

3.2. MIDI CONVERSION

To convert music file into a MIDI format, we need to find onsets, offsets and time duration of each key present in the music file. These parameters can be determined by using probabilistic methods to find musically relevant area in the music file. After finding the required parameters, we can form a piano-roll. Piano-roll is a video formatted audio file, in which it depicts fall of musical keys at certain time in a musical sequence. This piano-roll can be converted into MIDI file for further evaluation.

It is easy to extract parameters required for music transcription using MIDI file as it can be done by using simple python commands and pretty_midi library while programming the device.
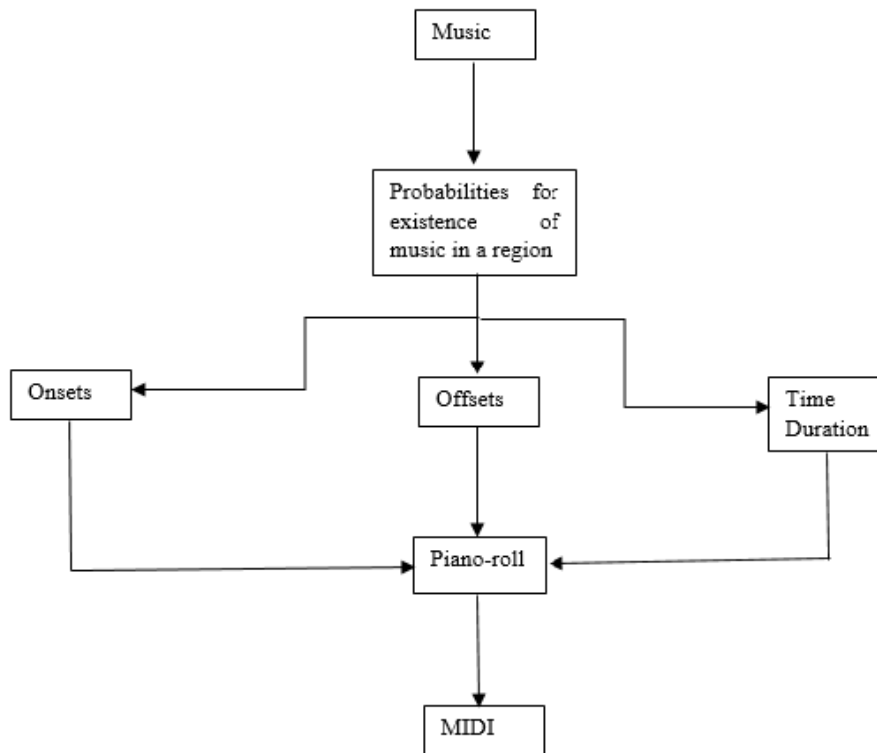

Fig. 4: Block Diagram for MIDI conversion

4. CONCLUSION

There are numerous ways for music transcription and this field is much challenging and emerging field, much more research has to be done. All the proposed techniques in this paper can be helpful in building an electronic device that can be used to transcribe the music. Music transcription can be further extended to image processing, where different symbols from a music sheet can be analysed or vice versa. Creating a music transcription system can help many people to learn music and encourage learners to move in an easy way. Besides an electronic device, a software application can be built to transcribe both online and offline music files.

5. REFERENCES

[1] J. J. Burred, M. Haller, S. Jin, A. Samour, and T. Sikora, "Audio content analysis," Semant. Multimed. Ontol. Theory Appl., pp. 123–162, 2008, DOI: 10.1007/978-1-84800-076-6_5.

[2] E. Benetos, S. Dixon, Z. Duan, and S. Ewert, "Automatic Music Transcription: An Overview," IEEE Signal Process. Mag., vol. 36, no. 1, pp. 20–30, 2019, DOI: 10.1109/MSP.2018.2869928.

[3]   M. Müller, D. P. W. Ellis, A. Klapuri, and G. Richard, "Signal processing for music analysis," IEEE J. Sel. Top. Signal Process., vol. 5, no. 6, pp. 1088–1110, 2011, DOI: 10.1109/JSTSP.2011.2112333.

[4]   A. Ycart and E. Benetos, "Learning and Evaluation Methodologies for Polyphonic Music Sequence Prediction with LSTMs," IEEE/ACM Trans. Audio Speech Lang. Process., vol. 28, pp. 1328–1341, 2020, DOI: 10.1109/TASLP.2020.2987130.

[5]   J. Sober-Mira, J. Calvo-Zaragoza, D. Rizo, and J. M. Inesta, "Pen-Based Music Document Transcription," Proc. Int. Conf. Doc. Anal. Recognition, ICDAR, vol. 2, pp. 21–22, 2018, doi: 10.1109/ICDAR.2017.258.

[6]   A. Grossman and J. Grossman, "Automatic Music Transcription : Generating MIDI From Audio," pp. 1–6, 2020.

[7]   A. Sadekar and S. P. Mahajan, "Polyphonic Piano Music Transcription using Long Short-Term Memory," 2019 10th Int. Conf. Comput. Commun. Netw. Technol. ICCCNT 2019, pp. 6–12, 2019, DOI: 10.1109/ICCCNT45670.2019.8944400.

[8]   K. Tanaka, T. Nakatsuka, R. Nishikimi, K. Yoshii, and S. Morishima, "Multi-instrument Music Transcription Based on Deep Spherical Clustering of Spectrograms and Pitchgrams," ISMIR, Int. Soc. Music Inf. Retr. Conf., pp. 327–334, 2020.

[9]   C. Dittmar and J. Abeßer, "Automatic Music Transcription with User Interaction," Digit. Media, pp. 567–568, 2008.

[10] X. Pan, B. Song, C. Liu, and H. Zhang, "Music detecting and recording system based on support vector machine," Proc. - 2019 Int. Conf. Commun. Inf. Syst. Comput. Eng. CISCE 2019, pp. 244–248, 2019, DOI: 10.1109/CISCE.2019.00063.

[11] A. McLeod and M. Steedman, "Evaluating automatic polyphonic music transcription," Proc. 19th Int. Soc. Music Inf. Retr. Conf. ISMIR 2018, pp. 42–49, 2018.

[12] A. Klapuri and T. Virtanen, "Automatic Music Transcription," Handb. Signal Process. Acoust., pp. 277–303, 2008, DOI: 10.1007/978-0-387-30441-0_20.