Bridging Sound and Symbol: A Music Transcription System

Desai Sumit Jitendra , Kamble Sanika Nagoji , Bhagat Neha Sunil , Bhate Udit Rajesh

Undergraduate Student, Computer Science & Engineering, D. Y. Patil Technical Campus Talsande, Kolhapur, Maharashtra, India

Undergraduate Student, Computer Science & Engineering, D Y Patil Technical Campus Talsande, Kolhapur, Maharashtra, India

Undergraduate Student, Computer Science & Engineering, D Y Patil Technical Campus Talsande, Kolhapur, Maharashtra, India

Undergraduate Student, Computer Science & Engineering, D Y Patil Technical Campus Talsande, Kolhapur, Maharashtra, India

ABSTRACT

The Automatic Music Transcription Mega Project aims to develop a comprehensive system for converting audio recordings of music into readable musical notation. Leveraging advanced techniques in machine learning, signal processing, and artificial intelligence, this project focuses on the accurate identification of musical elements such as pitch, rhythm, and timbre across various genres and instruments. The system will incorporate state-of-the-art algorithms for audio analysis, enabling real-time transcription and enhancing accessibility for musicians and educators. Through extensive dataset training and evaluation, the project seeks to address current limitations in transcription accuracy and scalability, ultimately creating a versatile tool for music composition, education, and preservation. The outcomes are expected to contribute significantly to both academic research and practical applications in the music industry.

Keyword *: Automatic Music Transcription, Audio to Sheet Music, Machine Learning in Music, Signal Processing, Real-Time Music Transcription, Pitch and Rhythm Detection, Music AI Systems, Musical Notation Generation.*

---

1 . INTRODUCTION

The Automatic Music Transcription (AMT) project aims to convert music audio into readable musical notes or sheet music using a computer. Just like a human musician listens to a song and writes down the notes being played, this system performs the same task automatically with the help of computer programs. This can be very helpful for musicians, music learners, and composers, as it saves time and effort in writing music manually. The main goal of this project is to build a system that can take an audio file as input, detect the musical notes and their timing, and then convert them into a readable format like sheet music or a MIDI file. This process makes it easier to understand and reuse music for different purposes such as learning, performance, or analysis.

2 . PROBLEM STATEMENT

The problem with traditional music transcription is that it is a time-consuming and complex process, requiring skilled musicians or transcribers to listen to an audio recording and manually write down the notes and rhythms. This becomes especially difficult with complex, multiinstrumental pieces or fast-paced music, where multiple notes overlap or instruments blend together. Many musicians, especially those without formal music training, struggle to transcribe their own compositions. In music education, teachers often spend valuable time transcribing pieces for students, limiting their ability to focus on other aspects of teaching. Additionally, researchers looking to study large volumes of music need an efficient way to convert audio recordings into written form. There is a clear need for an Automatic Music Transcription (AMT)system that can accurately and quickly convert music from audio to sheet music or digital formats, making the process more accessible and reducing the time and effort required for musicians, educators, and researchers.

3 . LITERATURE REVIEW

The field of music transcription has seen significant evolution over the past few decades, especially with advancements in digital signal processing, artificial intelligence, and machine learning. This section reviews existing research, systems, and technological developments that have contributed to the transformation of music transcription from manual efforts to automated solutions. It explores the progression from traditional rule- based approaches to modern machine learning-based models, evaluating their effectiveness in identifying musical elements such as pitch, rhythm, and timbre. The review highlights various tools, techniques, and datasets that have shaped the current landscape of automatic music transcription and discusses the challenges related to accuracy, real-time processing, and polyphonic complexity.

3.1 Early Approaches in Music Transcription Using Signal Processing

Initial developments in music transcription heavily relied on rule-based methods and signal processing techniques such as Fast Fourier Transform (FFT), zero-crossing rate, and autocorrelation. These methods were efficient for monophonic music but struggled with polyphonic textures. Tools like Sonic Visualiser and Audacity plugins enabled pitch tracking and note estimation through spectral analysis, making transcription more accessible for basic music structures.

3.2 Machine Learning for Automatic Music Transcription

With the rise of machine learning, transcription systems have become more accurate and adaptive. Supervised learning models trained on large datasets can now recognize complex musical elements and patterns across various instruments and genres. Models like “Onsets and Frames” by Google introduced a deep learning-based approach that significantly improved pitch and rhythm detection in polyphonic music. However, these systems require high computational resources and extensive training data.

3.3 Challenges in Music Transcription Systems

Despite advancements, several challenges persist in automatic transcription. Polyphonic music remains difficult due to overlapping harmonics and diverse instrument timbres. Real-time transcription also poses challenges in latency and accuracy. Additionally, the variability of musical styles and recording quality affects system performance. There is a need for robust, scalable, and interpretable solutions that balance accuracy and efficiency.

3.4 Evaluation Metrics and Benchmarking

Multiple datasets and metrics have been proposed to evaluate music transcription systems, such as the MAPS dataset and MIR-Eval metrics. These benchmarks allow for consistent performance comparison across different models. Research has also focused on note-level accuracy, onset detection precision, and pitch correctness as standard evaluation criteria for transcription systems.

4 . METHODOLOGY

This study is based on extensive secondary research and practical implementation of a music transcription tool using traditional audio signal processing and selected machine learning techniques. The goal was to design a system that effectively converts audio input into musical notation, especially for monophonic or simple polyphonic recordings.

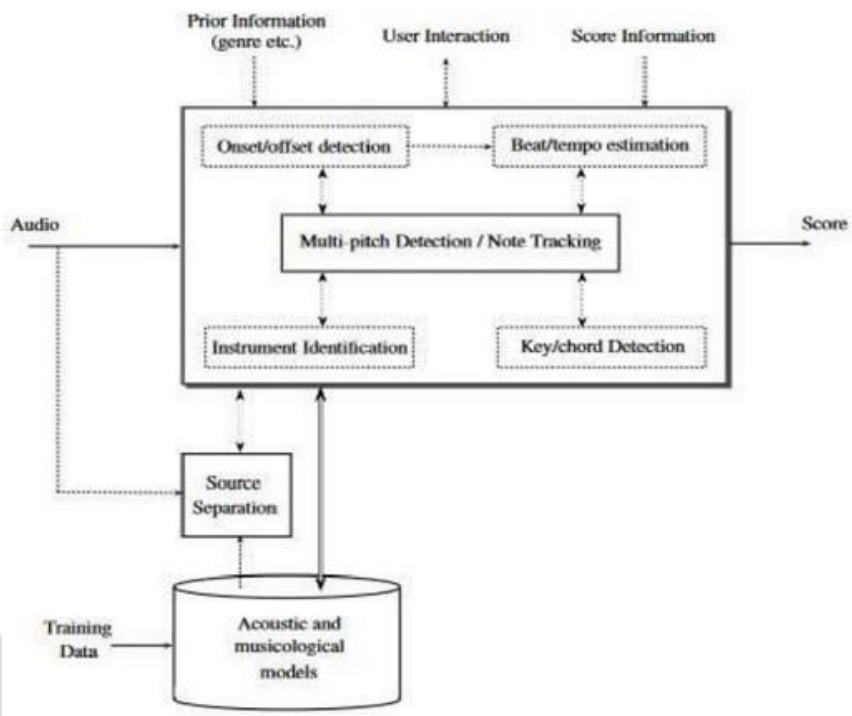

Fig 1 : System Architecture Diagram

The development process included the following stages:

- Audio Preprocessing: Audio signals are filtered and normalized to remove background noise
- and maintain consistency across samples.
- Feature Extraction: Core musical features such as pitch, onset, and duration are extracted
- using FFT, spectral centroid, and envelope detection.
- Note Mapping: Detected frequencies are mapped to corresponding musical notes based on
- standard pitch frequency charts.
- Rhythmic Analysis: Onsets are used to calculate note timing and duration, forming the basis
- of rhythmic transcription.
- Output Generation: The final transcription is output as a text file, visual waveform with note
- labels, or as MIDI for further editing.

The system was implemented using Python with libraries such as librosa, numpy, scipy, and matplotlib. The model was tested on instrument-based audio samples like piano, guitar and flute. Usability and accuracy were evaluated based on successful note prediction and rhythm alignment, ensuring the system can be used by musicians, learners, and researchers.

5. CONCLUSION

The Automatic Music Transcription project successfully demonstrates how machine learning and audio processing techniques can be used to convert music audio into readable musical notation. By using tools like Librosa, TensorFlow, and MIDI libraries, the system can detect pitch, rhythm, and note timing with reasonable accuracy. This project helps in bridging the gap between recorded audio and written music, making it easier for musicians, learners, and composers to understand and reuse music. Although transcription of complex polyphonic music remains a challenge, this system lays a strong foundation for future improvements. With access to better datasets and advanced models, the accuracy and flexibility of this system can be further enhanced. Overall, the project achieves its goal of creating a basic, working system that can automatically transcribe simple music audio files into musical notes, providing a valuable tool for music education and analysis.

6. REFERENCES

[1] Bonada, J., & Sillitti, R. (2014). Automatic Music Transcription: A Survey. This paper reviews different approaches to automatic music transcription, discussing methods for pitch detection, onset detection, and polyphonic music.

[2] McLeod, J., & Zafar, M. (2018). Deep Learning for Music Transcription. A study on how deep learning techniques, such as neural networks, can be applied to music transcription tasks, especially for polyphonic music.

[3] Müller, M. (2015). Fundamentals of Music Processing. This book provides an in-depth overview of the algorithms and tools used in music signal processing, including feature extraction and pitch detection.

[4] Hawkins, D., & Grill, T. (2020). Spleeter: A Fast and Efficient Music Source Separation Tool. This paper introduces the Spleeter tool for separating musical sources (e.g., vocals, drums, etc.), which is key in polyphonic transcription.

[5] Salamon, J., & Bello, J. P. (2012). A Tutorial on Onset Detection in Music Signals. A comprehensive guide on onset detection methods, which are crucial for determining when a note or sound starts in music.