K. J. SOMAIYA COLLEGE OF ENGINEERING

2UXC-601

DIGITAL SIGNAL PROCESSING

Submitted To:
Dr. Jagannath Nirmal
Head of Department
Department of Electronics
Prof. Bhargavi Kaslikar
Professor
Department of Electronics

Submitted By :
Meet Doshi
1812078
Annas Khan
1812080
Rohit Singh
1812083
Semester VI

CASE STUDY: MUSIC NOTES EXTRACTION

Meet Doshi[1], Annas Khan[2], Rohit Singh[3],
Department of Electronics Engineering
K.J. Somaiya College of Engineering, Vidyavihar, 400077, Maharashtra, India

I. ABSTRACT

Music signal processing may appear to be the junior relation of the large and mature field of speech signal processing, not least because many techniques and representations originally developed for speech have been applied to music, often with good results. However, music signals possess specific acoustic and structural characteristics that distinguish them from spoken language or other non-musical signals. This paper provides a signal analysis algorithm that specifically can figure out notes from its original file. We will examine how particular characteristics of music signals impact and determine these techniques, and we highlight a number of novel music analysis and retrieval tasks that such processing makes possible. Our goal is to demonstrate that, to be successful, music audio signal processing techniques must be informed by a deep and thorough insight into the nature of music itself.

II. INTRODUCTION

This chapter presents about what music is and what roles it plays in society. Further this chapter covers what audio processing is and what part it plays in processing of music notes. Later we outline the chapters about to come next.

Music is a ubiquitous and vital part of the lives of billions of people worldwide. Musical creations and performances are among the most complex and intricate of our cultural artifacts, and the emotional power of music can touch us in surprising and profound ways. Music spans an enormous range of forms and styles, from simple, unaccompanied folk songs, to orchestras and other large ensembles, to a minutely constructed piece of electronic music resulting from months of work in the studio.

The revolution in music distribution and storage brought about by personal digital technology has simultaneously fueled tremendous interest in and attention to the ways that information technology can be applied to this kind of content. From browsing personal collections, to discovering new artists, to managing and protecting the rights of music creators, computers are now deeply involved in almost every aspect of music consumption, which is not even to mention their vital role in much of today's music production.

Processing of this amount of music and audio files required sophisticated algorithms that can do computations in polynomial time. Parsing audio files to convert a musical tune to a playable instrument rhythm requires a sophisticated model and model parameters need to be application specific.

This case study concerns to investigate a methodology to make raw music extraction code to using sophisticated signal processing techniques to make the process less expensive in terms of time and processing power. We first take a look into a algorithm which can be used to convert any audio file to its respective musical notes. Then we show what methods can be used in this process to make the application faster and smoother whilst also preserving musical notes.

III. LITERATURE REVIEW

This chapter presents literature review of three research papers by *Müller*[1], *Mauch*[2], *Ewert*[3]. *Müller's* [1] paper focuses on techniques in signal processing for music analysis like source separation, mono and multipitch estimation, main melody extraction, etc. textitMauch's[2] paper focuses on estimation of chords using a bayesian network model. *Ewert's*[3] research was focused more towards estimating note intensities using a parametric spectrogram model.

Müller's [1] paper concerns the application of signal processing techniques to music signals, in particular to the problems of analyzing an existing music signal (such as piece in a collection) to extract a wide variety of information and descriptions that may be important for different kinds of applications. They first cover what music is made up of and how each component can be broken down to extract meaningful notes out of it. Then they have a discussion on how musical pitch in violin and piano can be calculated. They note that contemporary western music is based on the "equal tempered" scale, which, by a happy mathematical coincidence, allows the octave to be divided into twelve equal steps on a logarithmic axis while still (almost) preserving intervals corresponding to the most pleasant note combinations. The equal division makes each frequency 1.06 larger than its predecessor, an interval known as a semitone. They also show how ubiquity of simultaneous pitches, with coincident or near-coincident harmonics, is a major challenge in the automatic analysis of music audio. They show how

$time-frequency$, $time-chroma$ and $log-frequency$ representations of music files should be done. Also the objective of onset detection is to determine the physical starting times of notes or other musical events as they occur in a music recording. The general idea is to capture sudden changes in the music signal, which are typically caused by the onset of novel events. As a result, one obtains a so-called novelty curve, the peaks of which indicate onset candidates. The peaks of the novelty curve typically indicate the positions of note onsets. Therefore, to explicitly determine the positions of note onsets, one employs peak picking strategies based on fixed or adaptive thresholding. They also review techniques for musical instrument recognition in signals where only one instrument is playing at a time. Systems developed for this purpose typically employ the supervised classification paradigm.

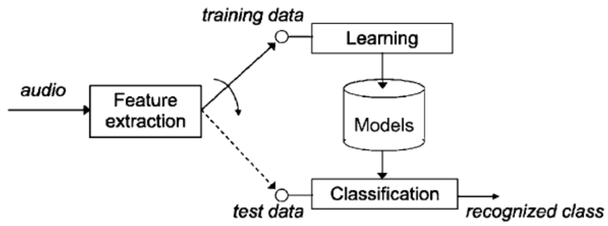

Fig. 1. General overview of supervised classification

The work described in this paper illustrates the broad range of sophisticated techniques that have been developed in the rapidly evolving field of music signal analysis, but as shown by this list of open questions, there is much room for improvement and for new inventions and discoveries, leading to more powerful and innovative applications. The aim of low-level processing in their case is to transform the audio input data into a representation which the high-level "musical" model can process. This representation consists of two different beat-synchronous chromagrams, one for the bass frequencies, and one for the treble frequencies, motivated by the importance of the bass note in harmony. They use techniques like note salience, tuning and chroma mapping, averaging beats and normalisation in their model to estimate chords.

Mauch's [2] underlying motivation of their research is to use automatic chord recognition to produce lead sheets where a lead sheet typically contains the melody written on traditional staves with time and key signature, along with chord symbols over the staves and the nominal bass note for the chord. They have presented a musically-informed dynamic Bayesian network for the automatic extraction of chord transcriptions from musical audio. The main novelty of this approach is simultaneous inference of metric position, key, chord, and bass pitch class, which reflects the natural interdependence of these entities. With 109 chord classes, their model provides a higher level of detail than previous approaches. Their network model topology, represented as a $2-slicetemporalBayesiannetworks$ (2-TBN) with two slices and six layers. The method presented achieves a mean correct overlap score of 71%. When compared ten different variants of this algorithm and show that each additional musical parameter significantly improves the method's performance. The greatest enhancement is achieved by additional bass modeling.

Ewert[3] presents automated methods for estimating note intensities in music recordings. Given a MIDI file (representing the score) and an audio recording (representing an interpretation) of a piece of music, their idea is to parametrize the spectrogram of the audio recording by exploiting the MIDI information and then to estimate the note intensities from the resulting model. The model is based on the idea of note-event spectrograms describing the part of a spectrogram that can be attributed to a given note event. After initializing this model with note events provided by the MIDI, they adapt all model parameters such that the model spectrogram approximates the audio spectrogram as accurately as possible. Given a MIDI file and an audio recording of a piece of music, their idea is to employ a parametric model that describes a spectrogram as a sum of note-event spectrograms. Here, each note-event spectrogram describes the part of a spectrogram that can be attributed to a specific note event. The approach starts by initializing the pitch, onset and duration parameters in the model using the note events provided by the MIDI file. In the second step, they adapt the onset and duration parameters by aligning the note events with their corresponding occurrences in the audio using a high-resolution music synchronization approach. In the third step, they iteratively modify parameters in their model related to the acoustic representation of a note event such that the model spectrogram approximates the audio spectrogram as accurately as possible. In a final step, the individual note intensities are estimated using the adapted note-event spectrograms described by the model. Their experiments on polyphonic piano music revealed measureable advantages of the method over a given baseline.

IV. METHODOLOGY

This chapter presents the methodology and dataflow through the model designed for recognition of musical notes. The flow chart describes what stages are present in processing the data. The algorithm shows exactly what techniques are used in each of these stages.

The basis for any audio processing application requires understanding of what the target audio will be made up of. For our application we know that the input is going to be a audio file which needs to be converted to singular notes, which then on playing back sound as close to the original audio as possible. Techniques like down sampling, FFT, averaging filters, moving average thresholds are used to achieve this.

A. FLOWCHART

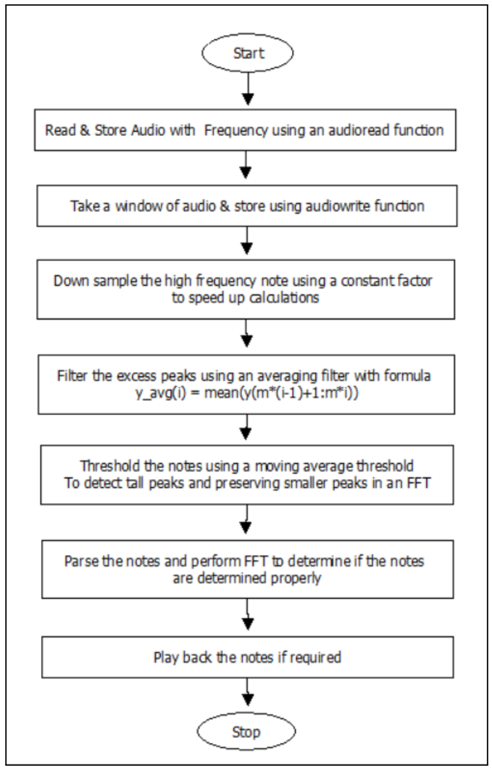

Fig. 2. Project Methodology Flowchart

B. ALGORITHM

Now we will define the exact algorithm which is used for our application. The application is run on Für Elise Bagatelle by Giscard Rasquin and Ludwig van Beethoven audio.

Steps:

1) Use MATLAB's audioread function as follows
[song,Fs] = audioread('Name of the file');
This will store the sampled audio file data in 'song' and the sample rate in Fs.
2) Speed up the song by a constant factor is the song is too slow
Fs=Fs*4;
3) Window a part of the song for analysing like this
y = song(t1:t2);
Now we have a sampled data of the real audio file from time t1 up to t2. Save the windowed audio file for future debugging.
4) Down sample the audio file to compute FFT and process audio faster
Fsm = round(Fs/20); Down Sampling factor 20.
5) Create a new vector yavg and initialize to 0's for filtering.
Use a for loop to use averaging filter
for i = 1:p
yavg(i) = mean(y(m*(i-1)+1:m*i));
Equation for the averaging filter

$$y[n] = \frac{1}{N} \sum_{i=0}^{N-1} x[n-i]$$

6) Calculate a threshold using the formula

$$thresh = 5 \times Median[|\ yavg(\max(1, i-5000) : i)\ |]$$

7) Threshold captures high frequency peaks whilst saving smaller important ones too.
if $yavg[i] > thresh$
for $j = 0 : 500$
if $i + j <= p$
$ythresh[i] = yavg[i]$
i = i+1
8) Create a function FreqToNote(f) which converts frequencies to their respective notes.
9) Now go through each note which was threshold and find their frequency using FreqToNote(f) function.
10) Play back each note and plot their FFT to make sure the notes are correctly mapped.

V. CONCLUSION AND FURTHER WORK

This chapter presents about what challenges are faced during music note extraction and covers what current works are in progress to eliminate those limitations. Furthermore it covers most promising research areas for analyzing music audio.

Signal processing for music analysis is a vibrant and rapidly evolving field of research, which can enrich the wider signal processing community with exciting applications and new problems. Music is arguably the most intricate and carefully constructed of any sound signal, and extracting information of relevance to listeners therefore requires the kinds of specialized methods that we have presented, able to take account of music specific characteristics including pitches, harmony, rhythm, and instrumentation. It is clear, however, that these techniques are not the end of the story for analyzing music audio, and many open questions and research areas remain to be more fully explored. Some of the most pressing and promising are listed below.

- Decomposing a complex music signal into different components can be a powerful preprocessing step for many applications. For example, since a cover song may preserve the original melody but alter the harmonization, or alternatively keep the same basic chord progression but devise a new melody (but rarely both), a promising approach to cover version recognition is to separate the main melody and accompaniment, and search in both parts independently.
- A different decomposition, into harmonic and percussive components, has brought benefits to tasks such as chord recognition, genre classification, and beat tracking. Note that such decomposition need not be perfect to yield benefits—even a modest improvement in the relative level between components can give a significant improvement in a subsequent analysis.
- Improved recognition and separation of sources in polyphonic audio remains a considerable challenge, with great potential to improve both music processing and much broader applications in computational auditory scene analysis and noise-robust speech recognition. In particular, the development, adaptation, and exploitation of sound source models for the purpose of source separation seems to be required in order to achieve an accuracy comparable to that of human listeners in dealing with polyphonic audio.
- In conjunction with the appropriate signal processing and representations, machine learning has had some great successes in music signal analysis. However, many areas are limited by the availability of high-quality labeled data. It is interesting to note that a given piece of music may have multiple, closely related sources of information, including alternate recordings or performances, partial mixes derived from the original studio multi tracks, score representations including MIDI versions, lyric transcriptions, etc. These different kinds of information, some available in large quantities, present opportunities for innovative processing that can solve otherwise intractable problems such as score-guided separation, generate substitutes for manual ground-truth labels using music synchronization techniques or use multi-perspective approaches to automatically evaluate algorithms.

REFERENCES

[1] Signal Processing for Music Analysis Meinard Müller, Member, IEEE, Daniel P. W. Ellis, Senior Member, IEEE, Anssi Klapuri, Member, IEEE, and Gaël Richard, Senior Member, IEEE
[2] Simultaneous Estimation of Chords and Musical Context From Audio Matthias Mauch, Student Member, IEEE, and Simon Dixon, Member, IEEE
[3] ESTIMATING NOTE INTENSITIES IN MUSIC RECORDINGS Sebastian Ewert University of Bonn Bonn, Germany Meinard Muller Saarland University and MPI Informatik Saarbrucken, Germany
[4] https://mypianonotes.com/fur-elise/
[5] https://in.mathworks.com/help/audio/index.html
[6] Identifying Missing and Extra Notes in Piano Recordings Using Score-Informed Dictionary Learning Siying Wang, Student Member, IEEE, Sebastian Ewert, Member, IEEE, and Simon Dixon
[7] Piano Transcription in the Studio Using an Extensible Alternating Directions Framework
[8] Music Conversion from Synthetic Piano to Chinese Guzheng Using Image-based Deep Learning Technique