Comparison of Repeating Pattern Extraction Techniques for Audio Pitch Detection

V.Sivaranjani, J.Umamaheswari

Abstract-Music separation methods are more demanding and complex, demanding system "training," user designation of special music features, and audio processing time to support their complicated frameworks. Pattern extraction from music strings is an complex problem. The repeated sequence extracted from music strings can be used as features for music extracted or compared. various works on music pattern extraction only focus on exact repeating patterns. However, music segments with minor differences may sound similar. Present the REpeating Pattern Extraction Technique (REPET), a novel and simple approach for separating the repeating "background" from the non-repeating "foreground" in a mixture. The basic idea is to identify the periodically repeating segments in the audio, compare them to a repeating segment model derived from them, and extract the repeating patterns via time-frequency masking. But in proposed system doesn't support the Small rhythmic patterns, but rhythmic patterns are essential for the balance of the music, and can be a way to identify a song. And enhanced a method to extract a monophonic rhythmic signature from a symbolic polyphonic score. To go beyond the simple extraction of all time intervals between onsets we select notes according to their length (short and long extractions) or their intensities (intensity+/− extractions). Once the frequency is calculated, now use dynamic programming to compare several sequences of audio.

Keywords- Pitch extraction, musical information retrieval, audio mining, pitch tracking, pattern extraction, audio segments.

I. INTRODUCTION

Due to large amount of data stored in web, the multimedia data explosion created an overload. This may cause information pollution problem because the user may not get the proper data or may get irrelevant data. In multimedia data like image, video etc, the audio plays an important role. Music information retrieval which is also called as Musical audio mining relates to the identification of perceptually important characteristics of a piece of music such as melodic, harmonic or rhythmic structure[8]. The popular audio search engines like musipedia. The recent research works on MIR systems include for Genre Classification, Emotion Classification, Composer Identification, Symbolic Classification, Query by Humming, Query by Example etc., But in case of Indian classical music, Raga is its basic component. Sentiments can be articulated through a particular raga, though the lyric or composition which has its own importance in our Indian classical music.

Manuscript Received May 2014

V. Sivaranjani, received her B.Tech Information technology in 2012 from Periyar Maniammai University, Thanjavur. Pursuing ME Computer and Science Engineering from Anna University, Chennai, Tamilnadu, India.

J. Umamaheswari, received her BE Computer Science and Engineering in 2012 from Anna University Chennai, Tamilnadu. Pursuing ME Computer and Communication Engineering from Anna University, Chennai, Tamilnadu, India.

These searches do not concentrate on audio feature. shazam [3] offer few ways for searching audio data based on the melodic contour, based on pitches and onset times, based on the rhythm alone or based on audio fingerprinting. Though they help in identifying pieces of music through various input methods, they
do not support raga based music retrieval.

II. LITERATURE SURVEY

Feature extraction refers to the process of extracting the distinguishing properties of an input data to classify them. An audio fingerprint is a compact content-based signature that summarizes an audio recording. Audio fingerprinting or content-based identification (CBID) technologies extract acoustic relevant characteristics of a piece of audio content and store them in a database. When presented with an unidentified piece of audio content, characteristics of that piece are calculated and matched against those stored in the database. Using fingerprints and matching algorithms, distorted versions of a single recording can be identified as the same music title [11].The individual notes or swaras are to be recognized as features in this module. An efficient way of recognizing swaras are explained in the works of Rajeswari Sridhar and T.V. Geetha [3] based on the different notes and their ratios of fundamental frequencies with frequency of the note "Sa". The final set swara combinations so found are mapped to their original audio files and can be forwarded for classification. In The work done by S. Shetty and K. K. Achary the features which are extracted from the input files are Swara Combination, Numbers of swaras used in the raga, Vakra pairs in Arohana & in Avarohana. Swara combination is the basic set of notes (swara) which is used to compose a musical script. In this paper the Swara combinations are represented in bits. This binary sequence is converted into decimal value [5].

III. PITCH DETECTION

A pitch detection(PD**)** is an algorithm designed to estimate the pitch or fundamental frequency of a quasi periodic or virtually periodic signal, usually a digital recording of speech or a musical note or tone. This can be done in the time domain or the frequency domain or both the two domains. PDAs are used in various contexts (e.g. phonetics, music information retrieval, speech coding, musical performance systems) and so there may be different demands placed upon the algorithm. If the signal is noise free and substantially a single tone, then just read the audio samples and count the zero crossings to determine frequency, remember two zero crossings per cycle. Take an FFT of the audio and for each time slice look for the bin with the highest energy. Use a set of band pass filters on the audio and pick the one with the highest energy[4].

The FFT approach is possibly the most common one, but they all have applications, it just depends on what you are trying to do. Automatic transcription of anything but the simplest monophonic music is basically a hard research problem in AI, but the simple stuff is not that difficult to code up

IV. SYSTEM DETAILS

The vocal and non-vocal regions by computing features such as MFCCs, Perceptual Linear Predictive coefficients (PLP), and Log Frequency Power Coefficients (LFPC), and using classifiers such as Neural Networks (NN) and Support Vector Machines (SVM). They then used Non-negative Matrix Factorization (NMF) to separate the spectrogram into vocal and non-vocal basic components. However, for an effective separation, NMF requires a proper initialization and the right number of components.

More recently, researchers in MIR have recognized the importance of repetition/similarity for music structure analysis. For visualizing the musical structure, Foote introduced the similarity matrix, a two-dimensional matrix where each bin measures the (dis)similarity between any two instances of the audio.

Durrieu et al. proposed to model a mixture as the sum of a signal of interest (lead) and a residual (background), where the background is parameterized as an unconstrained NMF model, and the lead as a source/filter model. They then separated the lead from the background by estimating the parameters of their model in an iterative way using an NMF-based framework. In addition, they incorporated a white noise spectrum in their decomposition to capture the unvoiced components.

Modules

- Identification of the repeating period
- Modeling of the repeating segment
- Extraction of the repeating patterns
- Frequency Based Extraction

Description

Identification of the repeating period

Repeating Period Identification Periodicities in a signal can be found by using the autocorrelation algorithm, which identify the similarity among a segment and a lagged version of itself over successive time intervals. Given a mixture signal x, we first calculate its Short-Time Fourier Transform (STFT) X, using half-overlapping Hamming windows of samples. We then derive the magnitude spectrogram by taking the absolute value of the elements of X, after discarding the symmetric part, while keeping the DC component. We then compute the autocorrelation of each row of the power spectrogram V^2(element-wise square of V) and obtain the matrix B .We use V^2 to emphasize the appearance of peaks of periodicity in B. If the mixture signal is stereo, is averaged over the channels. The overall acoustic self-similarity x of is obtained by taking the mean over the rows of B.

The idea is similar to the beat spectrum introduced except that no similarity matrix is explicitly calculated here and the dot product is used in lieu of the cosine similarity. Pilot experiments showed that this method allows for a clearer visualization of the beat structure in x . For simplicity, we will refer to as the beat spectrum .Once the beat spectrum is calculated, for the first term which calculate the pitch of the whole signal with itself (lag 0) is discarded. If repeating patterns are present in x,b would form peaks that are periodically repeating at different levels, revealing the underlying hierarchical repeating structure of the mixture.

Algorithm

Find repeating period p from beat spectrum b

l ← length of b after discarding the longest ¼ of lags

δ ← fixed deviation for possible shifted peaks

J ← empty array of length { $l/3$} for each possible period j in the first 1/3 of b do

$\Delta \leftarrow \left[\frac{3j}{4}\right]. I \leftarrow 0$ for each possible integer multiple I of j in b do $h \leftarrow \operatorname{argmax}_{k\in[i-\delta,i+\delta]} b(k)$

if $h \leftarrow \operatorname{argmax}_{k\in[i-\Delta,i+\Delta]} b(k)$ then $I \leftarrow I + b(h) - \operatorname{mean}_{k\in[i-\Delta,i+\Delta]} b(k)$

end if

end for

$p \leftarrow \operatorname{argmax}_{j} J(j)$

Modeling of the repeating segment

The rationale is that, assuming that the non-repeating foreground (≈ voice) has a sparse and varied time-frequency representation compared with the time-frequency representation of the repeating background ((≈ music) a reasonable assumption for voice in musi[2]c, time-frequency bins with little deviation at period would constitute a repeating pattern and would be captured by the median. Instead of, time-frequency beats with more deviations at period p would constitute a non-repeating pattern and would be removed by the median model. The median is preferred to the geometrical mean originally used because it was found to lead to a better discrimination between repeating and non-repeating patterns. Note that the use of the median is the reason why we chose to estimate the repeating period in the first 1/3 of the stable portion of the beat spectrum, because we need at least three segments to define a reasonable median. The segmentation of the mixture spectrogram V and the computation of the repeating segment model S.

Extraction of the repeating patterns

Once the repeating segment model S is calculated, we use it to derive a repeating spectrogram model W , by taking the element- wise minimum between and each of the segments of the spectrogram V, as exemplified in the bottom row .Non-negative spectrogram is the sum of a non-negative repeating spectrogram V and a non-negative non-repeating spectrogram V-M, then we must have $V \leq M$, element-wise, hence the use of the minimum function.

The time-frequency mask is then symmetrized and applied to the STFT of the mixture. The estimated music signal is obtained by inverting the resulting STFT domain[10]. The estimated voice signal is obtained by simply subtracting the time-domain music signal from the mixture signal x. The derivation of the repeating spectrogram model and the building of the soft time-frequency mask an M.

Frequency based Extraction

The simplest extraction is to consider all onsets of the song, reducing the polyphony to a simple combined monophonic track. This “notes on extraction” extracts durations from the inter-onset intervals of all consecutive groups of notes. For each note or each group of notes played simultaneously, the considered duration is the time interval between the onset of the current group of notes and the following onset. Each group of notes is taken into account and is represented in the extracted rhythmic pattern. However, such a notes on extraction is not really representative of the polyphony[7]: when several notes of different durations are played at the same time, there may be some notes that are more relevant than others.

Considering length of notes:

Focusing on the rhythm information, the first idea is to take into account the effective lengths of notes. At a given onset, for a note or a group of notes played simultaneously.

Considering intensity of onsets:

The second idea is to consider a filter on the number of notes at the same notes, extract only onsets with at least k notes (intensity+) or strictly less than k notes (intensity−), where the threshold k is chosen relative to the global intensity of the piece. The considered durations are then the time intervals between consecutive filtered groups.

V. CONCLUSION

Harmonic REpeating Pattern Extraction Technique (HREPET), a novel and simple approach for separating the repeating background from the non-repeating foreground in a mixture. The basic idea is to identify the randomly repeating segments in the audio, compare them to a repeating segment model derived from them, and extract the repeating patterns via time-frequency masking. REPET can be efficiently applied for music/voice separation, competing with two state-of-the-art approaches, while still showing room for improvement. More experiments on a data set real-world songs showed that REPET is robust to real-world recordings and can be easily extended to full-track songs. Further experiments showed that REPET can also be used as a preprocessor to pitch detection algorithms to improve melody extraction.

REFERENCES

1. Jonathan T. Foote, “An Overview of Audio Information Retrieval”, In Multimedia Systems, vol. 7 no. 1, pp. 2-11, ACM Press/Springer-Verlag, January 1999..
2. Kirthika, P. ; Chattamvelli, R. ,“A review of raga based music classification and music information retrieval (MIR)”, Engineering Education: Innovative Practices and Future Trends (AICERA), 2012 ,Digital Object Identifier: 10.1109/AICERA.2012.6306752 Publication Year: 2012 , Page(s): 1 – 5
3. Krishnaswamy, A, “Application of pitch tracking to South Indian classical music”, Applications of Signal Processing to Audio and Acoustics, 2003 IEEE Workshop on 19-22 Oct. 2003.
4. Lew, Michael S.; Sebe, Nicu; Djeraba, Chabane; Jain, Ramesh , “Content- based multimedia information retrieval: State of the art and challenges” , ACM Transactions on Multimedia Computing, Communications and Applications, Vol. 2, No. 1, February 2006, Pages 1–19..
5. Parag Chordia, “Automatic raag classification of pitch tracked performances using pitch-class and pitch-class dyad distributions”, In Proceedings of International Computer Music Conference, 2006.
6. https://ccrma.stanford.edu/~pdelac/research/MyPublishedPapers/icmc_2001-pitch_best.pdf
7. http://www.musipedia.org
8. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2629434/
9. http://www.riaa.org
10. http://www.shazam.com
11. http://en.wikipedia.org/wiki/Audio_mining and shazam

AUTHORS PROFILE

V. Sivaranjani received her B.Tech Information technology in 2012 from Periyar Maniammai University, Thanjavur. Pursuing ME Computer and Science Engineering from Anna University, Chennai, Tamilnadu. Area of interest includes Data Mining and Computer Network.

J. Umamaheswari received her BE Computer Science and Engineering in 2012 from Anna University Chennai, Tamilnadu. Pursuing ME Computer and Communication Engineering from Anna University, Chennai, Tamilnadu. Area of interest includes Data Mining, Computer Network and Network Security