IJCAT International Journal of Computing and Technology, Volume 1, Issue 3, April 2014
ISSN : 2348 - 6090
www.IJCAT.org

Various Pitch Extraction Techniques of Audio Files for Audio Information Retrieval

[1]V.Sivaranjani

PG Scholar, SNS College of Engineering, Coimbatore.

Abstract - Repetition is very important for the analysis of structure in music. To efficiently separate a audio into its music and voice components and group the musical and voice segments based on frequency and pitch a new method of automatically characterizing the rhythm and tempo of music and audio. The beat spectrum is a measure of acoustic self-similarity as a function of time lag. to visualizing the time structure of music and audio. The acoustic similarity between any two instants of an audio recording is calculated and displayed as a two-dimensional representation. Highly structured or repetitive music will have strong beat spectrum peaks at the repetition times. This reveals both tempo and the relative strength of particular beats, and therefore can distinguish between different kinds of rhythms at the same tempo. Unlike previous approaches to tempo analysis, the beat spectrum does not depend on particular attributes such as energy or frequency, and thus will work for any music or audio in any genre.

Keywords **- Pitch Extraction**

1. Introduction

Audio mining is a technique the content of an audio signal can be automatically analysed and searched for the given input.[5] To identified the periodically repeating patterns in the audio (eg: drum loop).the small rhythmic patterns does not support for the essential balance of the music and a way to extract a monophonic rhythmic signature in the music. That can be built from various different features such as similarity matrix, spectrogram, pitch contour and Chromagram. To be measure the dissimilarity and similarity between any two instances of the given audio. Anyone who has ever tapped a foot in time to music has performed rhythm analysis. Though simple for humans, this task is considerably more difficult to automate. We introduce a new measure of tempo analysis called the beat spectrum. This is a measure of acoustic self-similarity versus lag time, computed from a representation of spectrally similarity. Peaks in the beat spectrum correspond to major rhythmic components of the source audio. The repetition time of each component can be determined by the lag time of the corresponding peak, while the relative amplitudes of different peaks reflects the strengths of their corresponding rhythmic components. We also present the beat spectrogram which graphically illustrates rhythmic variation over time. The beat spectrogram is an image formed from the beat spectrum over successive windows. Strong rhythmic components are visible as bright bars in the beat spectrogram, making changes in tempo or time signature visible. In addition, a measure of audio novelty can be computed that measures how novel the source audio is at any time [7]. Instances when this measure is large correspond to significant audio changes. Periodic peaks correspond to rhythmic periodicity in the music. In the final section, we present various applications of the beat spectrum, including music retrieval by rhythmic similarity, an "automatic DJ" that can smoothly sequence music with similar tempos and automatic music video generation.

Properties of Musical Sounds
Rhythm

Defined as the particular arrangement of notes lengths in a piece of music. the flow of music through time[6]. the effect created by combining a variety of notes with different durations. Rhythm has several interrelated aspects: beat, tempo, meter.

BEAT

A regular, recurrent pulsation that divides music into equal unit of time. the beat of a rhythm is an essential feature of its personality.
Ex: when you clap your hands or tap your foot to music you are responding to its beat.

TEMPO

The speed of the music. the speed of the beat, that is the basic pace of the music. A tempo indication is usually given at the beginning of a piece. a fast tempo can be defined as associated with a feeling of energy and excitement in music. a slow tempo often contributes to a solemn and lyrical.

PITCH

Pitch is the relative highness or lowness that we hear in a sound and a pitch of a sound is determined by the frequency of its vibrations. two tones will sound different when they have different pitches .the distance in the pitch between any two tones called an interval. the faster vibrations and higher pitch. slower pitch vibration

,lower pitch. the frequency is measured in cycles per second.

2. Literature Survey

As digital audio collections grow in size and number, audio summarization, or "thumbnailing" has become an increasingly active research area.[2] Audio summaries are useful in applications such as e-commerce and information retrieval, because of the large file sizes and high bandwidth requirements of multimedia data. Quite often it is not practical to audit an entire work, for example if a music search engine returns many results each lasting several minutes. A representative excerpt that gives a good idea of the work is thus is desirable. Similarly, e-commerce music sites often make short song segments available to preview before purchase. In an audio retrieval system[9], it may make sense to judge the similarity of representative excerpts of a work rather than the work as a whole, especially if the analysis is computationally expensive. There is no point analyzing an entire symphony if a reasonable index can be derived from a ten-second excerpt.

This approach does not rely on semantic content that can't be automatically extracted, and thus cannot be considered optimal in that sense. For example, a summary of the first movement of Beethoven's Fifth Symphony without the famous four-note theme would not be ideal by most standards. To rectify this, the process can be weighted to reflect any available semantic information. Another possible desiderata for an audio summary is that it contain all representative portions[6]. For example, a popular song containing verses, refrains and a bridge should arguably be summarized by an example containing portions of all three segments. This is generally not possible with a short, contiguous excerpt. To find the most representative excerpt, we maximize the average segment similarity to the entire work. After window-based audio parameterization, a quantitative similarity measure is calculated between every pair of windows, and the results are embedded in a 2-D similarity matrix. Summing the similarity matrix over the support of a segment results in a measure of how similar that segment is to the whole.

Popular song containing verses, refrains and a bridge should arguably be summarized by an example containing portions of all three segments. This is generally not possible with a short, contiguous excerpt. Measure is maximized to find the segment that best represents the entire work. Then measure the variations on the method, and present experimental results for orchestral music, popular songs, and jazz.Contemporary content-based music retrieval applications have highlighted the need to extract rhythmic features from raw polyphonic audio recordings, in order to increase the efficiency of tools that perform a diversity of tasks, including musical genre classification, query-by-humming and query-by-rhythm,[3] to name but a few, e.g .Toward this end, several attempts have been made to create an algorithmic perception of rhythm. Most research has focused on tempo tracking, whereas, on the other hand, music meter extraction has attracted significantly less attention.

Method for the extraction of music meter and tempo from raw polyphonic audio recordings, assuming that music meter remains constant throughout the recoding. This assumption is acceptable for a large corpus of Greek traditional dance music, which has been in the center of our study. Our approach is based on the fact that the diagonals of the self-similarity matrix of the audio recording reveal periodicities corresponding to music meter and beat. By examining such periodicities it is possible to jointly estimate the music meter and tempo of the recording.

Extraction of music meter and tempo from raw polyphonic audio recordings, assuming that music meter remains constant throughout the recoding[4]. Although this assumption can be restrictive for certain musical genres, it is acceptable for a large corpus of folklore eastern music styles, including Greek traditional dance music. Our approach is based on the self- similarity analysis of the audio recording and does not assume the presence of percussive instruments. Its novelty lies in the fact that music meter and tempo are jointly determined.

Variety of feature candidates and their combinations, we chose to focus on two variations of the Mel-frequency cep-strum coefficients. SSM is symmetric around the main diagonal, in the sequel it suffices to focus on its lower triangle.

3. 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[8] 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.

4. 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[12] 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.

Modules

CEPSTRUM
FREQUENCY BASED EXTRACTION
HARMONIC PRODUCT SPECTRUM

Description

Cepstrum

A method for extracting the fundamental tone frequency, this process is performed by assuming that the audio signal f (t) is the result of the convolution of the impulse response of the vocal tract h (t) with the signal emitted by the glottis s (t), the operation can be seen as equation.

$$f(t)=h(t)*s(t)$$

This method aims to desconvulate signal f (t) equation and obtain s (t). To accomplish this we work with equation:

$$F(w)=H(w)*S(w)$$

To perform this procedure the first process is to decompose the real part of the imaginary part as shown below:

FFT(log$|F(w)|$)=FFT(log$|H(w)*S(w)|$)
=FFT(log$|H(w)|$+log$|S(w)|$)
=FFT(log$|H(w)|$)+FFT(log $|FF$T()|)

Therefore one can determine the Cepstrum as in equation.

$$C= F\text{FT}(\log\ |F(w)|)$$

Cepstrum is used in equation which determines the fundamental pitch.

$$f0=fs/(q-1)$$

This entire procedure can be seen as a mathematical diagram as shown in Figure 4, observing it is a Fourier Transformation and subsequently uses a natural logarithm and Inverse Transformation. Finally, a view allowing a higher peak can be the fundamental tone, but equation 6 is applied to determine more precision.

HPS (Harmonic Product Spectrum)

This is a method using the Fourier frequency transform on the set of short time segments fitted with a windowing function of the input signal. This data serves to emphasize the harmonics that are within the audio, making a record, and the values that occur most often is the fundamental tone. One of the advantages of this algorithm is that it is not necessary to know the first harmonic, but for subsequent comparison the fundamental tone for segment analysis can be determined.
This procedure is performed based on and algorithm using the following steps: Signal to be analyzed and the show window that applies to each. Reduce the signal by a factor of two; consequently locate the second fundamental frequency harmonic. Multiplying this previous segment fundamental frequency change often allows us to observe the fundamental frequency. This procedure is applied as often as necessary depending on the number of segments to be analyzed. Finally, it can be seen that from periodic frequency repeating the fundamental tone is determined.

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.

5. Performance Evaluation

Finally performance measures can then be defined: Source-to-Distortion Ratio (SDR), Source-to-Interferences Ratio (SIR) and Sources-to-Artifacts Ratio (SAR). To measure performance in pitch estimation, we used the precision, recall, and -measure. We define true positive (tp) to be the number of correctly estimated pitch values compared with the ground truth pitch contour, false positive (fp) the number of incorrectly estimated pitch values, and false negative (fn) the number of incorrectly estimated non-pitch values.
Precision (P) =tp/ (tp+fp)
Recall (R)=tp/(tp+fn)
F measure = 2(P*R)/(P+R)

$$SDR = 10log_{10}\left(\frac{||s_{target}||^2}{||e_{interf} + e_{artif}||^2}\right)$$

$$SIR = 10log_{10}\left(\frac{||s_{target}||^2}{||e_{interf}||^2}\right)$$

$$SAR = 10log_{10}\left(\frac{||s_{target} + e_{interf}||^2}{||e_{artif}||^2}\right)$$

6. Conclusion

In the music analysis from the beat spectrum that we extract the repeating period. In the system in order to improve the stability and performance of the system we propose a HPS and Cepstrum. Based on the structure used for the production of fundamental tone, it was found that the methods implemented, based on frequency domain techniques, have superior performance for monitoring the signal to noise extremes background. Some frequency domain algorithms require detailed mechanisms in adjustment to ensure optimum operating conditions so they can determine the fundamental tone. The HPS algorithm presents a satisfactory answer as to average values of the fundamental tone, while having acceptable characteristics in processing speeds. The only method that showed better stability in the frequency analysis was the HPS because the windowing application was performed for each segment showing more consistent value in determining the fundamental tone. We can obtain the efficient result than the existing system. Since the proposed system has more efficiency.

References

[1] S. Shetty and K. K. Achary. "Raga mining of Indian music by extracting arohana-avarohana pattern", International Journal of Recent Trends in Engineering,1(1), 2009.
[2] Foote, J., "Automatic Audio Segmentation using a Measure of Audio Novelty," in Proc. ICME 2000.
[3] Apte, Vasudeo Govind, "The Concise Sanskrit English Dictionary". Delhi: Motilal Banarsidas, 1987.
[4] Ram Avtar Vir, "Theory of Indian Music" , Pankaj Publications, New Delhi, 1999.
[5] Rajeswari Sridhar and T.V. Geetha," Raga Identification of Carnatic music for Music Information Retrieval", International Journal of Recent Trends in Engineering, Vol. 1, No. 1, May 2009.
[6] Segura-Bernal Gabriel,Alvarez-Cedillo J.Antonio, Herrera-Lozada J.Carlos, Hernandez-Bolaños Miguel , "Comparative F0 Algorithms Based on Frequency Analysis " Vol. 4, No. 4April 2013 ISSN 2079-8407 , Journal of Emerging Trends in Computing and Information Sciences.
[7] S. Shetty and K. K. Achary. "Raga mining of Indian music by extracting arohana-avarohana pattern", International Journal of Recent Trends in Engineering, 1(1), 2009.
[8] Schneck DJ, Berger DS., "The role of music in physiologic accommodation.", IEEE Eng Med Biol Mag. 1999 Mar-Apr;18(2):44-53.
[9]https://ccrma.stanford.edu/~pdelac/research/MyPublishedPapers/icmc_2001-pitch_best.pdf
[10]http://www.riaa.org
[11]http://www.new.dli.ernet.in/rawdatauploadupload/2005ab9_55.pdf
[12]http://www.kamalmusiccenter.com/category/indian-classical-music

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.