ISSN 2348 – 8034
Impact Factor- 3.155

GLOBAL JOURNAL OF ENGINEERING SCIENCE AND RESEARCHES

REPEATING PATTERN EXTRACTION TECHNIQUE (REPET):A METHOD FOR MUSIC/VIDEO SEPEARTION

Wakchaure Amol Jalindar*[1] and Rahul M.Mulajkar[2]
*[1]Student,E &TC Engineering, Jaihind College Of Engineering Kuran, Maharashtra,India.
[2]Asst.Prof,E &TC Engineering, Jaihind College Of Engineering Kuran, Maharashtra,India.

ABSTRACT

Repetition is a basic & most important principle used in music. Most of musical segments are characterized or identified by an underlying repeating structure over which continuously varying elements are superimposed. It is found real for pop songs where a singer always covers varying vocals on a repeating background. By using this base, we introduce the Repeating Pattern Extraction Technique (REPET), a single and non-complex method for isolating the repeating "background" from the non-repeating "foreground" in a mixture. The idea is to identify the periodically repeating segments in the audio, distinguish them to a repeating segment model obtained from them, and separate the repeating patterns by time-frequency screening. Experiments on 1,000 songs and 14 full-track real-world songs showed that this method can be successfully applied for music/voice separation, challenging with two recent state-of-the-art approaches. Other experiments showed that REPET can also be used as a preprocessor for pitch recognition algorithms to improve melody extraction technique.

Keywords: *Music classification, MFCC, Feature extraction, MMFS, Adaptive algorithms .*

I. INTRODUCTION

In this paper we concentrate on separating the singing voice signal from its musical background in audio. & will consider it as a special case of isolating out a human voice from structured background noise This task has some important practical applications, such as melody dictation from musical mixtures, removal of repetitive background noise for improved speech recognition, automatic karaoke defined the ability for the user to directly interact with the musical content of the songs. Also we can extend REPET by allowing the processing of long musical signals. Differences in the repeating background can be achieved by without proper segmentation of the audio. We also present a soft masking strategy, In this the TF mask is not binary. This type of postponement of REPET involves three main challenges. First, it depends on the calculations of the time-varying period of the repeating background. Second, it needs calculations of the corresponding locally periodic musical signal.by using this, it involves the descent of a TF mask to perform separation or isolation. Current methods in music/voice separation don't use the analysis of the repeating structure as a basis for isolation. We use different view to separating the main melody from the background accompaniment & find the repeating patterns in the audio and quotation them from the non-repeating elements.

II. MUSIC STRUCTURE ANALYSIS

Schenker declared that repetition is gives rise to the concept of the motive,defined as the smallest structural section within a musical piece. Ruwet defined repetition as a measure for partitioning music into small parts, see-through the arrangement of the musical piece. Ockelford recommended that repetition is what takes order to music, and order is what makes music artistically pleasing. Bartsch spotted choruses in general music by analyzing the structural termination in a similarity matrix built from the chromagram. Other audio methods include Cooper et al. who artificial a similarity matrix using MFCCs .Dannenberg et al. created a narrative of the musical structure related to the AABA form by monophonic pitch estimation, and probabilities associated with the model. In this paper, we first find useful information for classification from the symbolic demonstrations of music data. A similarity measure considering human perception of music is then designed to measure the similarity degree between two music objects. At last, we consider a broader coverage of music with seven classes to do performance evaluation.

III. MUSIC/VOICE SEPARATION

Music/voice separation methods identify the vocal/non-vocal segments, and then use a multiplicity of techniques to separate the lead vocals from the background ,it including spectrogram factorization, supplement model learning, and pitch-based inference techniques. Vembuet al. first identified the vocal and non-vocal sections by computing features such as MFCCs, Perceptual Linear Predictive coefficients (PLP), and Log Regularity Power Factors (LFPC), and using classifiers such as Neural Networks (NN) and Funding 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 operative separation, NMF requires a proper initialization and the right number of constituents. Raj et al. used a priori known non-vocal segments to train an enhancement model based on a Probabilistic Latent Component Analysis (PLCA). They then fixed the complement model to learn the vocal parts. Ozerovet al. first performed a vocal/non-vocal segmentation using MFCCs and Gaussian Mixture Models (GMM). They then trained Bayesian models to adapt an accompaniment model learned from the non-vocal segments. For an effective departure, such accompaniment model learning techniques require a sufficient amount of non-vocal segments and an accurate vocal/non-vocal prior cessation. Hsu et al. first used a Hidden Markov Model (HMM) to identify accessory, voiced, and unvoiced segments. They then used the pitch based decision method of Li et al. to separate the voiced vocals, while the pitch contour was imitative from the biggest pitch valuation algorithm of Dressler. In addition, they proposed a method to dispersed the unvoiced vocals based on GMMs and a method to enhance the voiced vocals based on spectral subtraction . This is a state-of-the-art system we compare to in our evaluation.

IV. GENERATION OF SIGNIFICANT REPEATING PATTERNS

Based on the above representations, a repeating pattern means a consecutive sequence that appears frequently in the recurrent or melodic sequence of a music piece. Hsu,Liu, and Chen [9] propose an algorithm for finding repeating patterns from a music piece. In this paper, we adapt this algorithm to the needs of music classification by since the following constraints:

a) Maximum length: Long sequences tend to contain duplicate information. The maximum limitation on the classification length will reduce matching information and the extra costs for pattern discovery.

b) Minimum length: Short sequences often have little material about the music and therefore its arrangement. The minimum constraint on the sequence length will alleviate the needless loads due to a large amount of short sequences.

c) Minimum frequency: The frequency of a sequence stands for the number of its occurrences in the music. The more frequency a sequence has in the music, the more demonstrative it will be. And the minimum constraint on frequency will diminish minor sequences to make the discovered patterns more significant.

V.MEL FREQUENCY CEPSTRAL COEFFICIENTS (MFCC)

MFCC coefficients model the spectral energy distribution in a perceptually evocative way. MFCCs are the most widely used acoustic feature for speech recognition, speaker recognition and Audio classification. MFCCs take into account certain properties of the Human auditory system.1) Critical band frequency resolution approx. 2)Log power in dB magnitudes.

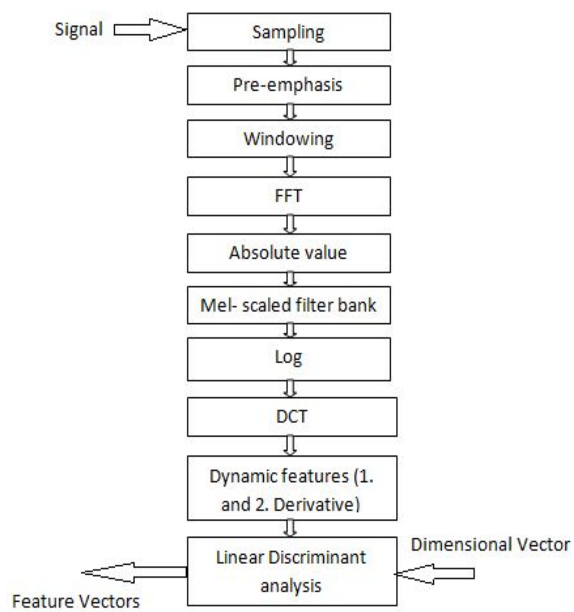

Fig1. Pipeline of MFCC

The Mel-frequency Cepstrum Coefficient (MFCC) technique is used to generate the mark of the sound collections. The MFCC are based on the known deviation of the human ear's critical bandwidth frequencies with filters spaced linearly at low frequencies and logarithmically at high frequencies used to deten the important characteristics of speech. Studies have shown that human insight of the frequency matters of sounds for speech signals does not follow a linear scale. Thus for each tone with an actual frequency, *f*, measured in Hz, a particular pitch is measured on a scale called the Mel scale. The Mel-frequency scale is rectilinear frequency spacing below 1000 Hz and a logarithmic spacing above 1000 Hz. As a hint point, the pitch of a 1 kHz tone, 40 dB above the perceptual hearing threshold, is defined as 1000 Mels. The following formula is used to compute the Mels for a exact frequency: mel(f) = 2595*log10(1+ f / 700).

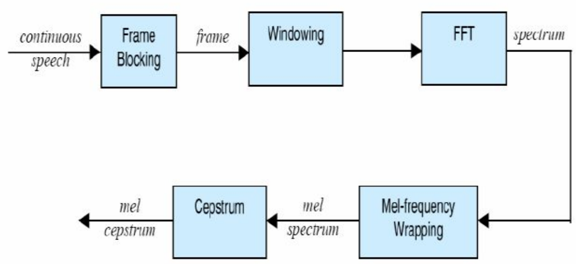

Fig 2.Block diagram of MFCC

The speech waveform is cropped to take away silence or acoustical intervention that may be present in the beginning or end of the sound file. The windowing block restrains the discontinuities of the signal by pointed the beginning and end of each frame to zero. The FFT block decodes each frame from the time domain to the frequency domain. In the Mel-frequency covering block, the indicator is plotted against the Mel scale to mimic human hearing. In the final step, the Cepstrum, the Mel-spectrum scale is converted back to ordinary frequency scale. This spectrum provides a good illustration of the spectral possessions of the signal which is key for representing and distinctive appearances of the speaker. After the fingerprint is created, we will also referred to as an acoustic vector. This vector will be stored as a place in the database. When an unknown sound file is imported into MatLab, a fingerprint will be created of it also and its resultant vector will be related against those in the catalog, again using the Euclidian distance technique, and a proper match will be unwavering. This process is as referred to as feature matching.

VI. RECOMMENDED RESEARCH

We propose an improved repeating period estimation algorithm, an enhanced repeating division modeling, and an alternate way for construction the time-frequency masking. We also proposition a simple procedure to extend the method stretched musical pieces. An algorithm to evaluation the repeating period had been developed and employed.

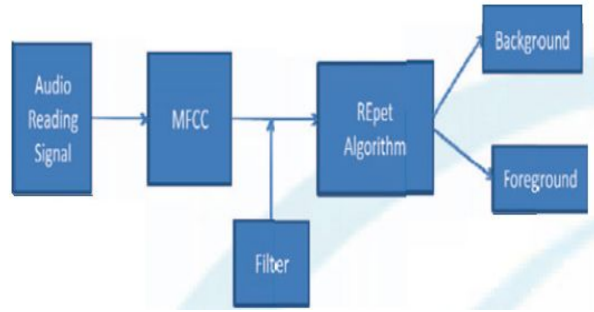

Fig 3. Block Diagram of Music/Voice separation.

VII. IDENTIFICATION OF THE REPEATING PERIOD

Repeating Period Identification Periodicities in a signal can be found by using the autocorrelation algorithm, to categorize the similarity among a sector and a lagged kind of itself over following time intervals. Given a blend signal x, we first calculate its Short-Time Fourier Transform (STFT) X, using half-overlapping Hamming windows of samples. We then derive the amount spectrogram by taking the absolute value of the elements of X, after clearance the symmetric part, while keeping the DC factor. We then compute the autocorrelation of each row of the power spectrogram (element-wise square of V) and obtain the matrix B .We use to underline the appearance of peaks of periodicity in B. If the mixture signal is stereo, is averaged over the channels. The overall audile self-similarity x of is obtained by taking the mean over the rows of B. The idea is like to the beat extent 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 meditation of the beat planning 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 rejected. If retelling forms are present in x,b would form peaks that are periodically repeating at different levels, tight-fitting the causal ordered repeating erection of the mixture.

VIII. FORMING OF THE REPEATING SEGMENT

The reasoning is that, assuming that the non-repeating foreground has a sparse and varied time-frequency representation compared with the time-frequency representation of the repeating background reasonable assumption for voice in music, time-frequency bins with little deviation at period would constitute a repeating pattern and would be captured

ISSN 2348 – 8034
Impact Factor- 3.155

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.

a)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 illustrated 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 , element-wise, hence the use of the smallest function. The time-frequency mask is then symmetrized and applied to the STFT of the mixture. The projected music signal is obtained by inverting the resulting STFT domain[10]. The estimated voice signal is obtained by simply extracting 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.

b) evaluation

Recently, FitzGerald et al proposed the Multipass Median Filtering based Separation (MMFS) method, a rather simple and novel tactic for music/voice separation. Their approach is based on a median filtering of the spectrogram at different frequency resolutions, in such a mode that the harmonic and percussive basics of the complement can be smoothed out, leaving out the vocals. To evaluate their method, they happily found recordings unrestricted by the pop band The Beach Boys, where some of the complete original accompaniments and vocals were made available as split stereo tracks 3 and separated tracks 4. After resynchronizing the embellishments and vocals for the latter case, we created a total of 14 sources in the form of split stereo wave files sampled at 44.1 kHz, with the complete garnish and vocals on the left and right channels, respectively. First, we compared the results of REPET with binary mask vs. soft mask, and without high-pass vs. with high-pass. A (non-parametric) Kruskal-Wallis one-way enquiry of discrepancy showed that using high-pass at 100 Hz on the voice estimations gave overall statistically better results, except for the voice SAR. Moreover, using a soft mask gave overall slightly better results, except for the voice SIR. The enhancement was however statistically not significant, except for the voice SAR. We nevertheless believe that the estimates sound perceptually better when using a soft mask instead of a binary mask, therefore we decided to show the results only for the soft mask. Since FitzGerald et al did not comment which tracks they use and only provided mean values, we could not conduct a statistical analysis to equate the results. We can however compare their means with our means and standard deviations, in the form of error bars. voice-to-music ratios of -6, 0, and 6 dB, without and with High-Pass at 100 Hz. The means and ordinary deviations of REPET are denoted by the error bars and the means of MMFS are represented by the crosses.

IX.CONCLUSIONS

In this paper, we propose a innovative method for ordering music data by contents. We correspondingly extract rhythm and melody from music data and adapt the methods of outcome repeating patterns to the desires of music classification. Given a music piece, we present a scheme for creating significant repeating patterns. A way to estimate the usefulness of SRP for arrangement is also proposed.The music to be classified, we incorporate human perception and musicology into the similarity measures for SRP matching. we provide a complete procedure for determining which class a music piece should be assigned to. The experiment results direct that some classes achieve better precision for a particular feature. Moreover, our method performs on average better than the HMM-based approach. Experiments on a data set of 1,000 song clips showed that 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 of 14 full-track 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. we have presented the Repeating Pattern Extraction Technique (REPET), a novel and simple approach for separating the repeating background from the non-repeating fore- ground in a mixture. The simple idea is to

ISSN 2348 – 8034
Impact Factor- 3.155

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.

X.ACKNOWLEDGEMENT

The authors would like to thank C.-L. Hsu for providing the results of his singing voice separation system, J.-L. Durrieu for helping with the code for his music/voice separation system, and A. Klapuri for providing the code for his multiple estimator. We also would like to thank A. Liutkus and his colleagues from Telecom Paristech for their prolific discussions, and our colleagues from the Interactive Audio Lab, M. Cartwright, Z. Duan, J. Han, and D. Little for their thoughtful comments. Finally, we would like to thank the reviewers for their helpful reviews.

REFERANCES

[1] R. Agrawal and R. Srikant, "Mining Sequential Patterns," Proceedings of IEEE Conference on Data Engineering, pp: 3-14, 1995.

[2] C. Anagnostopoulou and G. Westermann, "Classification in Music: A Computational Model for Paradigmatic Analysis," Proceedings of the International Computer Music Conference, 1997.

[3] J. J. Aucouturier and F. Pachet, "Music Similarity Measures: What's the Use?" Proceedings of International Symposium on Music Information Retrieval, 2002.

[4] W. Chai and B. Vercoe, "Folk Music Classification Using Hidden Markov Models," Proceedings of International Conference on Artificial Intelligence, 2001.

[5] C. C. Chen and Arbee L.P. Chen, "Query by Rhythm: An Approach for Song Retrieval in Music Database," Proceedings of IEEE Workshop on Research Issues in Data Engineering, pp: 139-146, 1998.

[6] R. B. Dannenberg, B. Thom, and D. Watson, "A Machine Learning Approach to Musical Style Recognition," Proceedings of International Computer Music Conference, 1997.

[7] S. Downie and M. Nelson, "Evaluation of a Simple and Effective Music Information Retrieval Method," Proceedings of ACM SIGIR Conference, pp: 73-80, 2000.

[8] A. Ghias, H. Logan, D. Chamberlin, and B.C. Smith, "Query by Humming: Music Information Retrieval in an Audio Database," Proceedings of ACM Conference on Multimedia, pp: 231-236, 1995.

[9] J. L. Hsu, C. C. Liu, and Arbee L.P. Chen, "Discovering Nontrivial Repeating Patterns in Music Data," Proceedings of IEEE Transactions on Multimedia, pp: 311-325, 2001.

[10] S. Moshe, Dynamic Programming, Marcel Dekker Inc., 1992. [11] J. Paulus and A. Klapuri, "Measuring the Similarity of Rhythmic Patterns," Proceedings of International Symposium on Music Information Retrieval, 2002. [12] J. Pei, J. W. Han, B. Mortazavi-Asi, and H. Pinto, "PrefixSpan: Mining Sequential Pat-
terns Efficiently by Prefix-Projected Pattern Growth," Proceedings of IEEE Conference on Data Engineering, 2001.

[13] D. Pogue and S. Speck, Classical Music for Dummies, IDG books worldwide Inc., 1999. [14] G. Tzanetakis, A. Ermolinskyi, and P. Cook, "Pitch Histograms in Audio and Symbolic Music Information Retrieval," Proceedings of International Symposium on Music Information Retrieval, 2002.