ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

A Novel Approach for Music/voice Separation Using REPET

Mr.P.Rama Koteswara Rao
Associate Professor & HOD,
Department of ECE,
Usha Rama College of
Engineering & Technology,
Telaprolu, AP, India.

K.Babu Rao, M.Tech
Associate Professor,
Department of ECE,
Usha Rama College of
Engineering & Technology,
Telaprolu, AP, India.

D.Vijaya Kumar
Associate Professor,
Department of ECE,
Usha Rama College of
Engineering & Technology,
Telaprolu, AP, India.

A.Ravindra
M.Tech (VLSI Embedded),
Department of ECE,
Usha Rama College of
Engineering & Technology,
Telaprolu, AP, India.

Abstract:

Repetition is a core principle in music. Many musical pieces are characterized by an underlying repeating structure over which varying elements are superimposed. This is especially true for pop songs where a singer often overlays varying vocals on a repeating accompaniment. On this basis, we 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. Experiments on data sets of 1,000 song clips and 14 full-track real-world songs showed that this method can be successfully applied for music/voice separation, competing with two recent state-of-the-art approaches. Further experiments showed that REPET can also be used as a preprocessor to pitch detection algorithms to improve melody extraction.After synthesizing the noise its being compressed.

Introduction:

In Music Information Retrieval (MIR), researchers used repetition/similarity mainly for audio segmentation and summarization, and sometimes for rhythm estimation (see Section I-A). In this work, we show that we can also use the analysis of the repeating structure in music for source separation. The ability to efficiently separate a song into its music and voice components would be of great interest for a wide range of applications, among others instrument/vocalist identification, pitch/melody extraction, audio post processing, and karaoke gaming. Existing methods in music/voice separation do not explicitly use the analysis of the repeating structure as a basis for separation (see Section I-B). We take a fundamentally different approach to separating the lead melody from the background accompaniment: find the repeating patterns in the audio and extract them from the non-repeating elements.

A. Music Structure Analysis:

In music theory, Schenker asserted that repetition is what gives rise to the concept of the motive, which is defined as the smallest structural element within a musical piece. Ruwet used repetition as a criterion for dividing music into small parts, revealing the syntax of the musical piece. Ockelford argued that repetition/imitation is what brings order to music, and order is what makes music aesthetically pleasing. Bartsch detected choruses in popular music by analyzing the structural redundancy in a similarity matrix built from the chromagram. Other audio thumbnailing methods include Cooper et al. who built a similarity matrix using MFCCs .Dannenberg et al. generated a description of the musical structure related to the AABA form by using similarity ma- trices built from a monophonic pitch estimation, and also the chromagram and a polyphonic transcription. Other music summarization methods include Peeters who built sim- ilarity matrices using MFCCs, the chromagram, and dynamic rhythmic features. Foote et al. developed the beat spectrum, a measure of acoustic self-similarity as a function of the time lag, by using a similarity matrix built from the spectrogram. Other beat estimation methods include Pikrakiset al. who built a similarity matrix using MFCCs. For a thorough review on music structure analysis.

B. Music/Voice Separation :

Music/voice separation methods typically first identify the vocal/non-vocal segments, and then use a variety of techniques to separate the lead vocals from the background accompainment, including spectrogram factorization, accompaniment model learning, and pitch-based inference techniques. Vembuet al. first identified the vocal and non-vocal regions by computing features such as MFCCs, Perceptual Linear Predictive coefficients (PLP), and Log Frequency Power Coeffi- cients (LFPC), and using classifiers such as Neural Networks (NN) and Support Vector Machines (SVM).

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

They then used Non-negative Matrix Factorization (NMF) to separate the spectrogram into vocal and non-vocal basic components. How- ever, for an effective separation, NMF requires a proper initialization and the right number of components. Raj et al. used a priori known non-vocal segments to train an accompaniment model based on a Probabilistic Latent Com- ponent Analysis (PLCA). They then fixed the accompaniment 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. However, for an effective separation, such accompaniment model learning techniques require a sufficient amount of non-vocal segments and an accurate vocal/non-vocal prior segmentation.Hsu et al. first used a Hidden Markov Model (HMM) to identify accompaniment, voiced, and unvoiced segments. They then used the pitch-based inference method of Li et al. to separate the voiced vocals, while the pitch contour was derived from the predominant pitch estimation algorithm of Dressler. In addition, they proposed a method to separate the unvoiced vo- cals based on GMMs and a method to enhance the voiced vo- cals based on spectral subtraction . This is a state-of-the-art system we compare to in our evaluation.

C. Proposed Method :

We present the REpeating Pattern Extraction Technique (REPET), a simple and novel approach for separating a re- peating background from a non-repeating foreground. The basic idea is to identify the periodically repeating segments, compare them to a repeating segment model, and extract the repeating patterns via time-frequency masking . The justification for this approach is that many musical pieces can be understood as a repeating background over which a lead is superimposed that does not exhibit any immediate repeating structure. For excerpts with a relatively stable repeating back- ground, we show that REPET can be successfully applied for music/voice separation. For full-track songs, the repeating background is likely to show variations over time (e.g., verse followed by chorus). We therefore also propose a simple procedure to extend the method to longer musical pieces, by applying REPET on local windows of the signal over time (see Section V). Unlike other separation approaches, REPET does not depend on particular statistics (e.g., MFCC or chroma features), does not rely on complex frameworks (e.g., pitch-based inference techniques or source/filter modeling), and does not require preprocessing (e.g., vocal/non-vocal segmentation or prior training). Because it is only based on self-similarity, it has the advantage of being simple, fast, and blind. It is therefore, completely and easily automatable. A parallel can be drawn between REPET and background subtraction. Background subtraction is the process of separating a background scene from foreground objects in a sequence of video frames. The basic idea is the same, but the approaches are different. In background subtraction, no period estimation nor temporal segmentation are needed since the video frames already form a periodic sample.

Also, the variations of the back- ground have to be handled in a different manner since they involve characteristics typical of images. For a review on back- ground subtraction. REPET bears some similarity with the drum sound recognizer of Yoshii et al.. Their method iteratively updates time-frequency templates corresponding to drum patterns in the spectrogram, by taking the element-wise median of the patterns that are similar to a template, until convergence. As a comparison, REPET directly derives a whole repeating segment model by taking the element-wise median of all the periodically repeating segments in the spectrogram (see Section II).

Although REPET was defined here as a method for separating the repeating background from the non-repeating fore- ground in a musical mixture, it could be generalized to any kind of repeating patterns. In particular, it could be used in Active Noise Control (ANC) for removing periodic interferences. Ap- plications include canceling periodic interferences in electro- cardiography (e.g., the power-line interference), or in speech signals (e.g., a pilot communicating by radio from an aircraft). While REPET can be applied for periodic interferences removal, ANC algorithms cannot be applied for music/voice separation due to the simplicity of the models used.

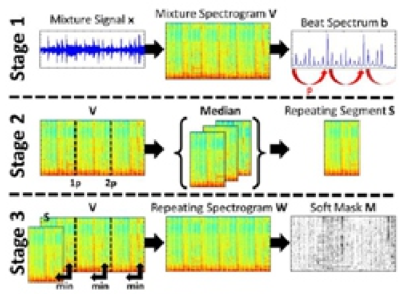


Fig. 1. Overview of the REPET algorithm. Stage 1: calculation of the beat spectrum and estimation of the repeating period . Stage 2: segmentation of the mixture spectrogram and computation of the repeating segment model . Stage 3: derivation of the repeating spectrogram model and building of the soft time-frequency mask .

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Melody extraction:

In this section, we evaluate REPET as a preprocessor for two pitch detection algorithms to improve melody extraction. We first introduce the two pitch detection algorithms(Section VI-A). We then present the performance measures (Section VI-B). We finally show the extraction results (Section VI-C).

A. Pitch Detection Algorithms:

We have shown that REPET can be successfully applied for music/voice separation. We now show that REPET can consequently improve melody extraction, by using it to first sepa- rate the repeating background, and then applying a pitch detection algorithm on the voice estimate to extract the pitch con- tour.

We employ two different pitch detection algorithms: the well-known single fundamental frequency (F0) estimator YIN proposed by de Cheveigné et al. in [35], and the more recent multipleestimator proposed by Klapuri in [36].

B. Performance Measures:

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. A pitch estimate was treated as correct if the absolute difference from the ground truth was less than 1 semitone.

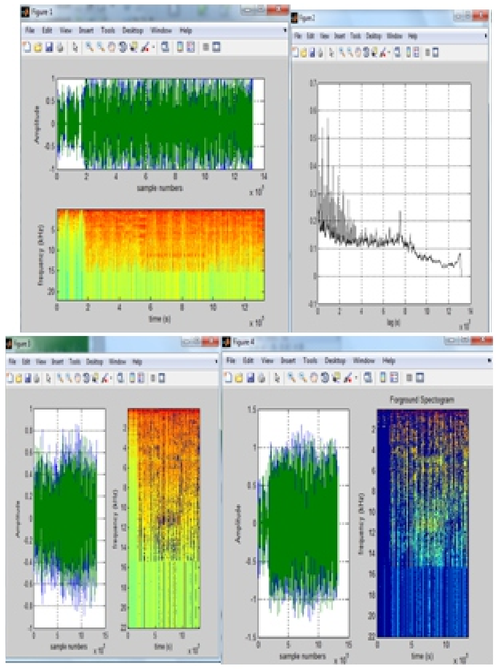
C. Extraction Results:

We extracted the pitch contours from the voice estimates obtained from REPET, including the potential enhancements (see Section IV-D), using YIN and Klapuri's system.

We also extracted the pitch contours from the mixtures and the voice sources to serve, respectively, as a lower-bound and upper-bound on the performance in pitch estimation. Perfor- mance in pitch estimation was measured by using the precision, recall, and -measure, in comparison with the ground truth pitch contours.

Fig:Melody extraction performance via the -measure, at voice-to-music ratios of (left column), 0 (middle column), and 5 dB (right column), using YIN (top plot) and Klapuri's system (bottom plot), on the mixtures (mixtures), on the voice estimates of REPET plus high-pass filtering (R + H), then enhanced with the best repeating period and the indices of the vocal frames (R+H+P+V), and on the voice sources (voices).

RESULTS:



ACKNOWLEDGMENT:

Conclusion:

In this work, we have presented 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. 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.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Further experiments showed that REPET can also be used as a preprocessor to pitch detection algorithms to improve melody extraction.After emphasizing the data and synthesizing the noise it is being compressed.

References:

[1] H. Schenker, Harmony. Chicago, IL: Univ. of Chicago Press, 1954.

[2] N. Ruwet and M. Everist, "Methods of analysis in musicology," Music Anal., vol. 6, no. 1/2, pp. 3–9+11–36, Mar.-Jul. 1987.

[3] A. Ockelford, Repetition in Music: Theoretical and MetatheoreticalPerspectives. Farnham, U.K.: Ashgate, 2005, vol. 13, Royal MusicalAssociation Monographs.

[4] J. Foote, "Visualizing music and audio using self-similarity," in Proc. 7th ACMInt. Conf. Multimedia (Part 1), Orlando, FL, Oct.-Nov. 30–05,1999, pp. 77–80.

[5] M. Cooper and J. Foote, "Automatic music summarization via similarity analysis," in Proc. 3rd Int. Conf. Music Inf. Retrieval, Paris, France, Oct. 13–17, 2002, pp. 81–85.

[6] A. Pikrakis, I. Antonopoulos, and S. Theodoridis, "Music meter and tempo tracking from raw polyphonic audio," in Proc. 9th Int. Conf.Music Inf. Retrieval, Barcelona, Spain, Oct. 10–14, 2008.

[7] G. Peeters, "Deriving musical structures from signal analysis for music audio summary generation: "Sequence" and "state" approach," in ComputerMusic Modeling and Retrieval,U.Wiil, Ed. Berlin/Heidelberg, Germany: Springer, 2004, vol. 2771, Lecture Notes in Computer Science, pp. 169–185.

[8] J. Foote, "Automatic audio segmentation using a measure of audionovelty," in Proc. IEEE Int. Conf. Multimedia and Expo, New York, Jul.-Aug. 30–02, 2000, vol. 1, pp. 452–455.

[9] J. Foote and S. Uchihashi, "The beat spectrum: A new approach to rhythm analysis," in Proc. IEEE Int. Conf. Multimedia and Expo, Tokyo, Japan, Aug. 22–25, 2001, pp. 881–884.

[10] M. A. Bartsch, "To catch a chorus using chroma-based representations for audio thumbnailing," in Proc. IEEE Workshop Applicat. SignalProcess. Audio Acoust., New Paltz, NY, Oct. 21–24, 2001, pp. 15–18.

Author's Details:

Rama Koteswara Rao

He is currently working as Associate Professor in ECE Department, Usha Rama Engineering and Technology, Telaprolu, AP, India. He is working towards his Ph.D. at Andhra University College of Engineering, Visakhapatnam, AP, India. He received his M.Tech. from the JNTU college of Engineering Anantapur. He has thirteen years of experience in teaching undergraduate students and three years industrial experience. His research interests are in the areas of speech signal processing, embedded systems and speaker recognition techniques. He is a life member of ISTE and ISOI.

K.Babu Rao
M.Tech, Associate Professor, Department of ECE,
Usha Rama College of Engineering & Technology,
Telaprolu, AP, India.

D.Vijaya Kumar
Associate Professor,Department of ECE,
Usha Rama College ofEngineering & Technology,
Telaprolu, AP, India.

A.Ravindra
M.Tech (VLSI Embedded),Department of ECE,
Usha Rama College ofEngineering & Technology,
Telaprolu, AP, India.