Survey on: Separation techniques for separation of Singing Voice from Music Accompaniment

Ms. Monali R. Pimpale
Department of Computer Engineering
Mumbai University, India
{monalipimpale20@gmail.com}

Prof. Shanthi Therese
Department of Computer Engineering
Mumbai University, India
{shanthitherese123@gmail.com}

Prof. VinayakShinde
Department of Computer Engineering
Mumbai University, India
{vdshinde@gmail.com}

Abstract—today the multimedia databases are growing rapidly on large scale. For the effective management and exploration of huge amount of music data the technology of singer identification is developed. Songs performed by particular singer can be clustered automatically with the help of this technology. To improve the performance of singer identification the technologies are revealed that can separate the singing voice from music accompaniment. These methods of separation are widely used in film industry for the purpose of song recording. This survey consists of different sound source separating methods.

*****

I. INTRODUCTION

The development of singer identification enables the effective management and exploration of huge amounts of music data. With the help of singer identification technology, songs performed by a particular singer can be automatically clustered for easy management or searching. There are many algorithms exists which are used for singer identification which are based on the concept of feature extraction which identifies the appropriate singer from the obtained features. In most popular music, singer's voice is combined with music accompaniment. Thereforethose methods extract features directly from the accompanied vocal segments, hence it's very difficult to acquire good performance when accompaniment is stronger or singing voice is weaker. To get better performance the techniques are revealed which separates the singing voice from music accompaniment. There are many sound source separation algorithms which separates the singing voice from music accompaniment. Sound source separation means the tasks of evaluating and identifying the signal produced by an individual sound source from a mixture signal which is consisting of multiple sources. This is a very fundamental or basic problem in many audio signal processing tasks, since analysis and processing of isolated sources can be done with very much better accuracy than the processing of mixtures signal. This paper studies the different sound source separation, method which is going separate vocal and non-vocal part of the song.

II..DIFFERENT METHODS FOR SOUND SOURCE SEPARATION

A. Repeating Pattern Extraction Technique (REPET):

Repetition is a core principle in popular music. Many musical pieces are characterized by a repeating structure over which varying elements are overlapped. This is especially true for pop songs where a singer often overlays varying vocals on a repeating accompaniment. This methodis based on this repeating accompaniment; the repeating "background" (e.g., a guitar riff or a drum loop and many more other instrumental music) is separated from the non-repeating "foreground" in a mixture. The fundamental idea behind this method is it identifies the periodically repeating segments in the audio then compares them to a repeating segment model derived from them, and extracts the repeating patterns via time-frequency masking. REPET Algorithm involves three stages: Identification of the repeating period, modeling of the repeating segment, Extraction of the repeating patterns[1],[2].

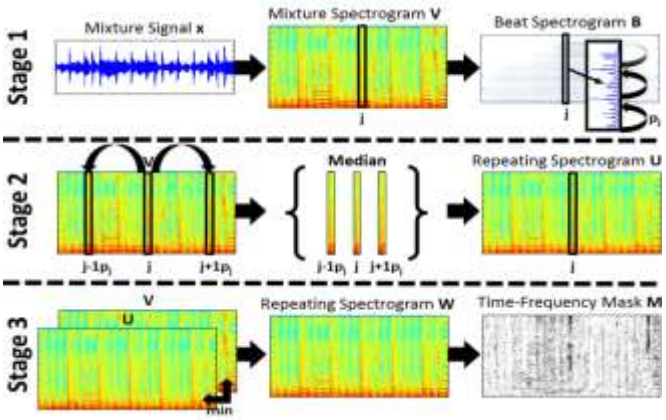


Fig.1. Overview of the REPET algorithm.

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.

B. Butterworth Low Pass Filter

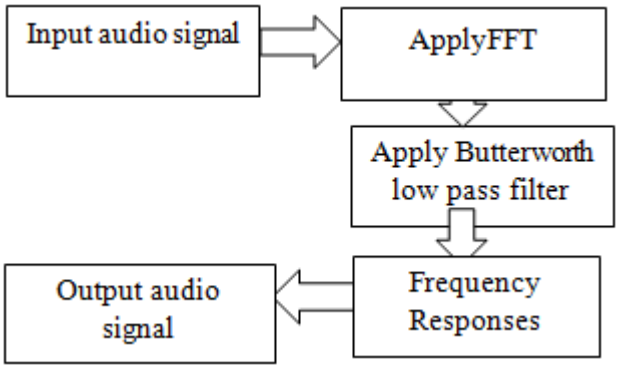


Fig 2.Block diagram of Butterworth low pass filter separation Method

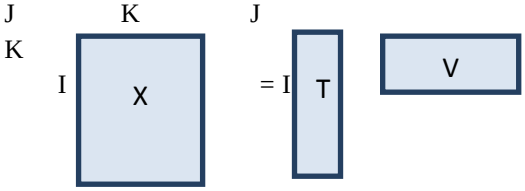
In Butterworth low pass filter separation method, low pass Butterworth filter is used for separating singing voice and music from audio file. In this method audio input signal is first converted into frequency domain using FFT and then Butterworth low pass filtering is applied. This method is based on proper selection of cut off frequency. Generally, for separation of singing voice from music the cut off frequency is taken as 3400 Hz. Hence as an output signal. The result of this method is not as accurate as other methods[7].

C.Non-negative Matrix Factorization (NMF)

Non-negative Matrix Factorization (NMF) is the separation method which is more than 30 year old. . NMF has been used in various applications, including image processing, brain computer interface, document clustering, collaborative predictions, and many more. Non-negative includes sparsity. Short-time Fourier Transform (STFT)is used to obtain complex value representation in the frequency domain. NMF imposes nonnegative constrain which lead only additive combination of original data. NMF can use long-window and short-window spectrogram factorization, it can give better performance for removing music interferences from singing voice . NMF works as a decomposition method. NMF used to decompose the mixture spectrogram into set of component to different sound sources[4].

Mathematical representation can be written as:

$$X_{ij} = T_{ki} V_{kj} \quad (1)$$



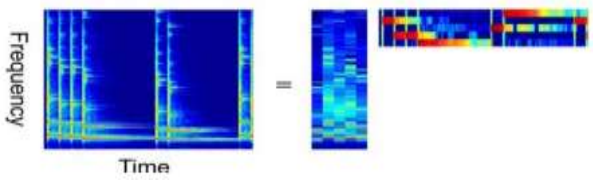


Fig.1 formulation of NMF (top) and its application to music signal (bottom)

The fig.1 shows decomposition of a given non-negative matrix X into two different smaller non-negative matrices T and V by using non negative matrix factorization. Observation matrix x is a phase-invariant time-frequency (T-F) representation (e.g. magnitude spectrogram or power spectrogram) of the input sound mixture, where is I the number of frequency channel and J is number of time frames[6].

D. Robust Principal Component Analysis (RPCA)

The input audio signal which is superimposition of singing voice and background music. Audio signal can be considered in term of data matrix which is combination of low rank components and sparse components, i.e. musical accompaniment and singing voice respectively as shown in fig.In first step compute Short-Time Fourier Transform (STFT) of the audio signal where signal is represented in time frequency domain. In the separation method, STFT of the input audio signal is calculated with the help of overlapping hamming window.After applying STFT, RPCA is applied by means of Augmented Langrange Multiplier (ALM) which performs the computational problem of RPCA. After applying RPCA data matrix producestwo output matrices “L” and “S” low rank matrix and sparse matrix respectively. Binary frequency mask is later applied for quality of separation result. Then Inverse Short Time Fourier transform (ISTFT) is applied, in order to obtain the waveform of the estimated results followed by evaluation of the results[10][12].

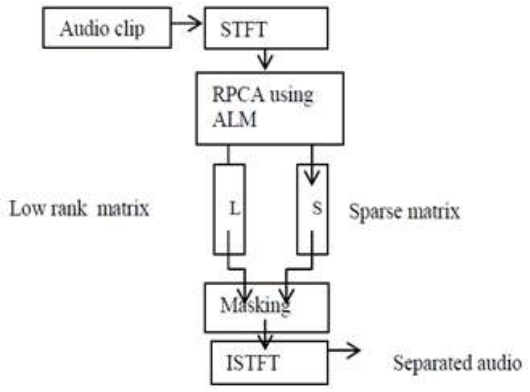


Fig. Robust principal component analysis

E. Non negative matrix partial co factorization:

Many algorithms based on the non-negative matrix factorization (NMF) were developed in applications for blind or semi-blind source separation and those NMF algorithms are efficient and robust for source separation when sources are statistically dependent under conditions that additional constraints are imposed such as non-negativity, sparsity, smoothness, lower complexity or better predictability. However, without any prior knowledge of a source signal, the standard NMF cannot separate specific source signal from the mixing signal. To tackle this problem, nonnegative matrix partial co-factorization (NMPCF) was introduced. NMPCF is a joint matrix decomposition integrating prior knowledge of singing voice and accompaniment, to separate the mixture signal into singing voice portion and accompaniment portion. Matrix co-factorizations can be served as a useful tool when side information matrices are available, in addition to the target matrix to be factorized. NMPCF was emerged from the concept of joint decomposition or collective matrix factorization, which make the multiple input matrices be decomposed into several factor matrices while some of them are shared, therefore, shows a greater potential in singing voice separation from monaural recordings [8].

Reference No.	**Singing Voice Separation Techniques**	**Performance Comparison**
[1],[2], [3],[4]	**Repeating Pattern Extraction Technique (REPET)**	Steps-1. Identification of the repeating period 2.Modeling of the repeating segment 3. Extraction of the repeating patterns. Disadvantage – □ Required to find Repeating structure of music
[7]	**Butterworth Low Pass Filter**	Steps- 1.ApplyFFT 2. Apply Butterworth low pass filter 3.Frequency Responses Disadvantage- gives less accuracy
[4], [6]	**Non-negative Matrix Factorization (NMF)**	Steps:1.Decomposition using NMF 2.Separation 3. Reconstruction Disadvantage – □ Imposes only the non-negativity constraint.
[10], [11], [12], [13], [13],[14]	**Robust Principal Component Analysis (RPCA)**	Steps – □ STFT □ RPCA □ ISTFT Advantages – □ Training data not required. □ Easy to understand. □ Many researchers used for separation. □ This method can achieve higher GNSDR.
[8],[9]	Non negative matrix partial co factorization	Steps: 1.Prepare the spectrogram of the target music signal X and the spectrogram of the prior signal Y Initialize factor matrices with random positive values. 2. Iterate Update each factor matrix using 3. Reconstruct the separated signals Advantage: this methods gives the better performance than other techniques

III. Conclusion

Separation of singing voice from music accompaniments is a branch of speech separation process or audio processing, which popular and interesting research topic for many years, but still it lacking behind in separating the vocal signal from

the mixture of signal with 100%accuracy. Many researcher have used many singing voice separation and singing voice classifier technique to separate singing voice from music accompaniments. A more efficient method is required for separation vocal and non- vocal part of a mixed signal.

REFERENCE

[1] Emad M. Grais, Ibrahim SayginTopkaya, HakanErdogan. “Audio Visual Speech Recognition Using Single Channel Source Separation”. Faculty of Engineering and Natural Sciences Sabanci University, Orhanli,Tuzla,34956,Istanbul.fgrais,isaygint,haerdogang@sabanciun iv.edu

[2] ZafarRafii, Bryan Pardo. “REpeating Pattern Extraction Technique (REPET): A Simple Method for Music/Voice Separation”.IEEE Transactions on Audio, Speech, and Language Processing, Vol. 21, No. 1, January 2013.

[3] ZafarRafii and Bryan Pardo, “A simple music/voice separation method based on the extraction of the repeating musical structure”, ICASSP, May 2011, pp.221-224.

[4] TuomasVirtanen ,”Unsupervised Learning Methods for Source Separation in Monaural Music Signals” Tuomas Virtanen

[5] Non-negative matrix factorization based compensation of music for automatic speech recognition, Bhiksha Raj, T. Virtanen, SourishChaudhure, Rita Singh, 2010.

[6] Siddhi, Desai, and Desai Bhavik. "Survey on Separation Methods of Music and Speech."

[7] J. Yoo et al., “Nonnegative matrix partial co-factorization for drum source separation,” in Proc. IEEE Int. Conf. Acoust. Speech, Signal Process., 2010, pp. 1942 1945.

[8] M. Kim et al., “Nonnegative matrix partial co-factorization for spectral and temporal drum source separation,” IEEE J. Sel. Topics Signal Process., vol. 5, no. 6, pp. 1192–1204, Dec. 2011.

[9] Po-.Sen. Huang, Scott Deeann Chen, Paris Smaragdis and Mark Hasegawa-Johnson, “Singing voice separation from monaural recordings using robust principal component analysis”, ICASSP, 2012.

[10] Zhouchen Lin, Minming Chen, Leqin Wu and Yi Ma, “The augmented Langrange multiplier method exact recovery of corrupted low-rank matrices”, Tech. Rep.UILU-ENG-09- 2215, UIUC, Nov.2009.

[11] Emmanuel J. Candes, Xiaodong Li, Yi Ma and John Wright, “Robust principal component analysis ?”, Journal of the ACM, vol.58, no. 3, article 11, pp.11:1-11:37, May 2011.

[12] John Wright, YigangPeng, Yi Ma, Arvind Ganesh and Shankar Rao, “Robust Principal Component Analysis: Exact Recovery of Corrupted Low-Rank Matrices by Convex Optimization”, pp. 1-

[13] Umap, Priyanka K., and Kirti B. Chaudhari. "Performance Analysis of RPCA Algorithm for Segregation of Singing Voice from Polyphonic Music."