Segmentation of Singing Voice and Musical Accompaniment in Audio Mixture Signal

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—we propose a system to separates singing voice from music accompaniment for monaural recordings to improve the performance of singer identification. Our system consists of two stages. The first stage implements the non-negative matrix partial co-factorization (NMPCF), to separate the mixture signal into singing voice portion and accompaniment portion. The second stage is based on the separated singing voice obtained by the first stage, pitches of singing voice are first estimated and then the harmonic components of singing voice can be distinguished. Spectrum of separated singing voice may contain extraneous frequency components Missing Feature Method can used to refine the separated signal of singing voice. This refined signal then proceeded for the stage of singer identification. The system will analyze the performance of singer identification based on song with accompaniment and pure singing voice.

*****

I. INTRODUCTION

Human auditory system has capability of separating sounds from different sources. However, separation of singing voice is quite natural to human auditory systems but the automated system to solve this problem is extremely difficult. Singing voice separation is quite challenging because of the absence of spatial information. There are many existing algorithms utilize the harmonic structure of singing voice and extract the singing pitch from the given input mixture for separation. There are systems that have addressed the problem of separating singing voice from music accompaniment systematically. A singing voice separation system has used in many areas such as automatic lyrics recognition. Automatic lyrics recognition requires the input to the system as solo singing voice, which is often unrealistic since for almost all songs, singing voice is accompanied by musical instruments. This requirement for such a system can be satisfied by successful separation of singing voice from music accompaniment. Separation of singing voice from music accompaniment is also used for the singer Identification which enables the effective management of large amount of music data. Using this technology, songs performed by a particular singer can be automatically clustered for easy management and searching. Most of the studies in the field of singer identification pay attention to features extraction directly from the songs. In popular music, singing voice is often interwoven with accompaniment i.e. music. So those methods based on the features extracted directly from the songs are difficult to achieve good performance when accompaniment is stronger or singing voice is weaker. There are some methods for reducing or removing the negative influences caused by instrumental accompaniment. In this paper we propose a novel method to identify the singer by separating vocal and non-vocal parts of the song by using non-negative matrix partial co factorization. The proposed model consist of two stages: separation of singing voice from music accompaniment using non negative matrix partial factorization and second stage is singer identification based on first stage.

II. RELATED WORK

Singer identification has a long and rich tradition of research. Over the years many new approaches have been devised in the field of singer identification. To separate the singing voice from music accompaniment for better performance Sound source separation methods are available. This section takes a look at some of these methods pertaining to separation of singing voice from music accompaniment and singer identification Non negative matrix partial co factorization for drum source separation is method which incorporates prior knowledge of drum, which does not require training sets related to pitched instruments. It is a method where target matrix and drum only matrix were jointly decomposed; NMPCF gives superior result than NMF+SVM [1].

A novel algorithm for separating sources from polyphonic accompaniment has been proposed. The method combines two approaches, pitch-based inference and unsupervised non-negative matrix factorization. NMF find outs the noise from vocal segments. Algorithm has produced clearly better results than the reference separation algorithms [2].

The proposed method improved NMPCF based predecessors by combining their heterogeneous assumption about spectraland temporal domains .The main contribution of this proposed method is that it provides adequate formation of input signal by grouping them into prior knowledge and portioned column block of mixture signal which are allocated to NMPCF learning process[3].

The algorithm which is on temporal continuity criteria which improve the pitch detection accuracy of pitched sounds. The framework of NMF produces better result than independent analysis [4].

The systematic approach to identify and separate the unvoiced singing voice from accompaniment. The method follows CASA which consist of segmentation and grouping Stage. The performance shows that quality of the separated voiced and unvoiced parts is improved [5].

Unsupervised method which is extension of NMF for extracting rhythmic sources from polyphonic music mixture. The method proposed here has used NMPCF which uses temporally crepitating sources across input mixture. The defined method has some aspect to improvement in area of optimal number of iterations and segments. It shows acceptable result but not superior quality than referred drum source separation system [6].

The system for separation of singing voice from music accompaniment for monaural recording consists of three stages. The singing voice detection stage partitions and classifies an input into vocal and non-vocal portions. For vocal portions, the predominant pitch detection stage detects the pitch of the singing voice and then the separation stage uses the detected pitch to group the time-frequency segments of the singing voice [9].

The approach for source separation method based on excitation and filter models of sound. Excitation corresponds to pitch information and filter corresponds to response of an instrument. The paper proposed excitation maximization algorithm which jointly filter responses and organize the excitation to filter [10]

The two methods, accompaniment sound reduction and reliable frame selection are proposed. It first extracts the harmonic components of the predominant melody from sound mixtures and then resynthesizes the melody by using a sinusoidal model driven by these components. The latter method then estimates the reliability of frame of the obtained melody by using two Gaussian mixture models (GMMs) for vocal and non-vocal frames to select the reliable vocal portions of musical pieces [11].

III. NONNEGATIVE MATRIX PARTIAL CO-FACTORIZATION FOR SINGING VOICE SEPARATION

There are many algorithms for sound source separation that are efficient and robust for source separation., when sources are statistically dependent under certain conditions that additional constraints are imposed such as non-negativity, lower complexity or better predictability , smoothness, sparsity. Without any prior knowledge of a source signal, the non negative matrix factorization can’t separate specific source signals from the mixing signal. To solve this problem, nonnegative matrix partial co-factorization (NMPCF) was introduced. The authors [1] separated drum sources from the mixture without any prior knowledge of drum sources. As we know most of the drum sources have the temporal property of repeatability. [7] Exploited the prior knowledge of drum sources to separate drum sources from the music mixtures.Along with the mixture signal to be separated the solo signal of various drums was used as auxiliary input for separation.

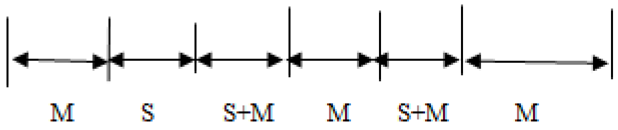
Here we proposed a method for separating singing voice from monaural songs, which are consist of solo singing voice and music accompaniment. If the spectrogram matrix of mixture signal is denoted as A then each element $A_{ft}$is represented by two attributes frequency *f* at at time *t*. If we apply standard NMF on this mixture nonnegative matrix A can be written as

$$A=SM$$

Matrix S represents the frequency values of thevarious sources contained in the mixture signal, and the corresponding rows of M represent the activations of frequency bases across the time. If we consider S matrix some values in it represents frequencies of singing voice and some are music frequencies. If we find the frequency basis vectors S_f representing the singing source and is respective activations St, then we can reconstruct the magnitude spectrogram of solo singing voice as $As=S_fSt$ However, the components representing singing are placed in the arbitrary locations of S , in order to distinguish the components of target source, the NMPCF was introduced. NMPCF conducts partial co-factorization simultaneously exploiting the prior knowledge of singing voice and accompaniment; it can locate the frequency bases of each source. So that the single-source signals can be separated from the mixing signals. By applying NMPCF on the mixture matrix A can be decomposed as

$$A=S_fS_t+ M_f M_t$$

In above expression$S_f$and M_f represents the frequency characteristics of the singing voice music respectively, and $S_t$and M_t represent the time characteristics of singing voice and music accompaniment.



Here S represents singing voice,

M represents Musical accompaniment,

S+M represent the singing voice with background music

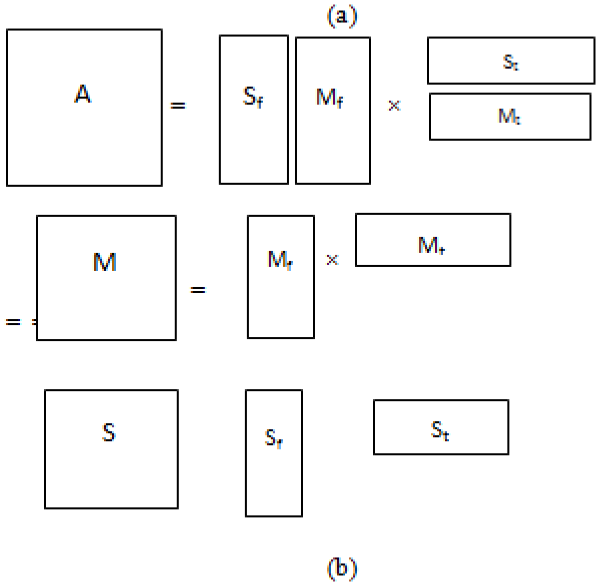


Fig 1: (a) decomposition of audio signal (b) Matrix factorization by NMPCF, matrix A, M, S represents mixture spectrogram, music accompaniment spectrogram and Singing voice spectrogram.

Fig1. Shows the factorization process of the NMCF. Here the mixture signal is consisting of the singing voice and music accompaniment. Generally vocal parts of the song are interludes between non vocal parts.

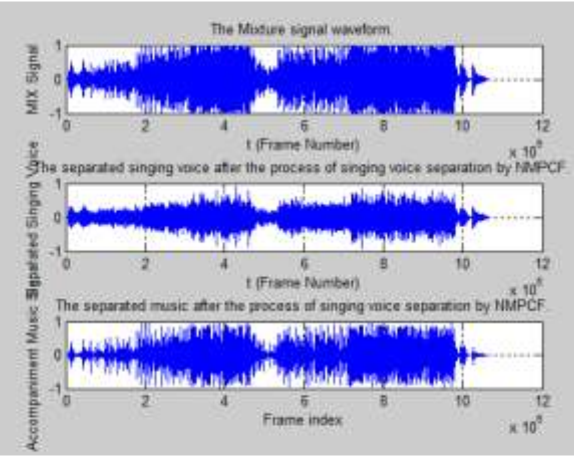


Fig: Spectrogram of mixed signal, Separated signal and accompaniment signal

Fig shows the spectrogram after implementing NMPCF. It shows the three spectrograms first spectrograms shows mixed signal, second spectrogram shows separated signal and third spectrogram shows the accompaniment

IV. SINGER IDENTIFICATION

Previous section describes the separation of singing voice from music accompaniment improving the performance of singer identification. In this section, we briefly describe a singer identification system. The singer identification consists of two stages. The first stage performs singing voice segregation based on CASA which first detects the pitches of singing voice directly over the song mixed by accompaniment then perform CASA-based segregation and finally reconstruct a complete spectrum of singing voice using missing feature method. Here, we perform the singer identification over the separated singing voice.

V. CONCLUSION

This paper presents a method to separate singing voice from music accompaniment. This separated singing voice signal can be further used for the singer identification as well as the music directors can use this clean signal to perform different operations various musical backgrounds.

REFERENCES

[1] 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

[2] T. Virtanen, A. Mesaros, and M. Ryynanen, “Combining pitch-based inference and non negative spectrogram factorization in separating vocals from polyphonic music,” in Proc.ISCA Tutorial Res. Workshop Statist.Percept.Audit.(SAPA), 2008.

[3] 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.

[4] T. Virtanen, “Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria,” IEEE Trans. Audio, Speech, Lang. Process., vol. 15, no. 3, pp. 1066–1074, Mar. 2007

[5] C. L. Hsu and J. Jang, “On the improvement of singing voice separation for monaural recordings using the MIR-1 K dataset,” IEEE Trans. Audio, Speech, Lang. Process, vol. 18, no. 2, pp. 310–319, Feb. 2010.

[6] M. Kim et al., “Blind rhythmic source separation: Nonnegativity and repeatability,” in Proc. IEEE Int. Conf. Acoust. Speech, Signal Process., 2010, pp. 2006–2009

[7] B. Raj, P. Smaragdis, M. Shashanka, and R. Singh, “Separating a foreground singer from background music,” in Proc. Int. Symp. Frontiers Res. Speech Music, Mysore, India, 2007.

[8] M. Ryynanen, T. Virtanen, J. Paulus, and A. Klapuri, “Accompaniment separation and karaoke application based on automatic melody transcription,” in Proc. IEEE Int. Conf. Multimedia Expo, 2008, pp. 1417–1420.

[9] Y. Li and D. L. Wang, “Separation of singing voice from music accompaniment for monaural recordings,” IEEE Trans. Audio, Speech, Lang. Process., vol. 15, no. 4, pp. 1475–1487, May 2007.

[10] A. Klapuri, T. Virtanen, and T. Heittola, “Sound source separation in monaural music signals using excitation-filter model and emalgorithm,”in Proc. IEEE Int. Conf. Acoust. Speech, Signal Process., 2010, pp. 5510–5513

[11] H. Fujihara, M. Goto, T. Kitahara, and H. G. Okuno, “A modeling of singing voice robust to accompaniment sounds and its application to singer identification and vocal-timbre-similarity-based music information retrieval,” IEEE Trans. Audio, Speech, Lang. Process, vol. 18, no. 3, pp. 638–648, Mar. 2010.