International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

Tandem Algorithm with Supervised Classifier for Pitch Estimation and Voice Separation from Music Accompaniments: Survey

Vikas Nichal[1], Mane V.A[2], Gadhave D.D[3]

[1]Shivaji University, Kolhapur. ADCET Ashta

[2]Anasaheb Dange college of engg.. Shivaji University, Kolhapur

[3]Shivaji University, Kolhapur. ADCET Ashta

Abstract: ***Singing pitch estimation and singing voice separation are challenging due to the presence of music accompaniments that are often non stationary and harmonic. The proposed work Singer voice extraction from music accompaniments have been found as a challenging task because of lack of efficient methods in recognizing the singer's voice. Singer voice recognition methods have faced many challenges, because of the attempts to extract it from music accompaniments. The proposed system uses tandem algorithm that estimates the singing pitch and separates the singing voice jointly and iteratively. To enhance the performance of the tandem algorithm for dealing with musical recordings, we propose a trend estimation algorithm to detect the pitch ranges of a singing voice in each time frame and Support vector machine (SVM) for pitch estimation and voice extraction from music accompaniments. It has proposed its advantages in voice and music processing.***

Keywords: iterative procedure, pitch extraction, singing voice separation, tandem algorithm.

1. Introduction

The music databases both with professional and personal requirements have rapidly grown because of popularization and wide usage of digital music. The trending of technologies that deal with categorization and retrieval has also risen in response to the requirements and consumer demands. The automatic singer voice extraction technology not only acts as an application, but also working with various applications and acts as sub-processes. The necessity of such technology has been extended to a wide end. This technology intends to extract a particular singer's voice from music accompaniments based on certain feature sets like pitch.

Pitch, can be often referred as fundamental frequency, is one of the important parameters in speech analysis, synthesis and vocoder applications. Hence, pitch can be considered as a potential feature to detect the voice signal. Many researchers have put forth significant attempts to devise pitch extraction methods. However, it is not a simple task to extract pitch from speech signals from a noisy environment or speech signal with accompanies. In many applications, the quality of the outcome is directly proportional to the accuracy of the pitch extraction method. Under such circumstances, if the pitch extraction is not effectual with noisy environment or accompanied speech signals, the entire quality of the system will get degraded. Correlation based methods are proved to be robust against noise, and to further improve the accuracy of the pitch extraction. However, singing voice separation is considered to be a special case of voice separation application and so it requires special attention towards pitch extraction methods.

The proposed work Singer voice extraction from music accompaniments have been found as a challenging task because of lack of efficient methods in recognizing the singer's voice. Singer voice recognition methods have faced many challenges, because of the attempts to extract it from music accompaniments. The proposed system uses tandem algorithm and Support vector machine (SVM) for pitch estimation and voice extraction from music accompaniments. It has proposed its advantages in voice and music processing.

2. Relevance

There are different traditional methods developed for singing voice separation from Music like .The singing voice separation by using a harmonic-locked loop technique. In this system, the fundamental frequency of the singing voice needs to be known a priori. The system also does not distinguish singing voice from other musical sounds. When the singing voice is absent the system incorrectly tracks partials that belong to some other harmonic source. The harmonic-locked loop requires the estimation of a partials instantaneous frequency, which is not reliable in the presence of other partials and other sound sources. Therefore, the system only works in conditions where the energy ratio of singing voice to accompaniments is high [7].

The Separation of singing and piano sound ; this system requires significant amount of prior knowledge , such as the partial tracks of premixing singing voice and piano or the music score for piano sound. This prior knowledge in most cases is not available. Therefore the system cannot be applied for most real recording [6].

The Monaural speech segregation technique based on pitch tracking and amplitude modulation. This system relies heavily on pitch to group segments. Therefore the accuracy of pitch detection is critical. This system obtains its initial

pitch estimation from the time lag corresponding to the maximum of a summary of autocorrelation function. This estimation of pitch is unreliable for singing voice. This system assumes that voice speech is always present. For singing voice separation, this assumption is not valid. This system cannot separate unvoiced speech [5]

Another method Computational auditory scene analysis (CASA) is proposed in this method a lot of effort has been made to segregate speech from music accompaniments. But the performance of current CASA system is still limited by pitch estimation errors and residual noise [4].

The tandem algorithm used in voice speech segregation performs pitch estimation and voice separation jointly and iteratively. It is observed that the target pitch can be estimated from a few harmonics of the target signal. On the other hand, it can separate some target signals without perfect pitch estimation. This system show consistent performance improvement for all types of intrusion except rock music, presumably because of the strong harmonicity of the music Accompaniment. This indicates that separating speech from music is challenging to their tandem algorithm [3].

In the proposed work tandem algorithm is used to reduce pitch detection problem by using a trend estimation algorithm to bound the singing pitch contours in a series of time-frequency (T-F) blocks that have much narrower pitch ranges as compared to the entire possible range. The estimated trend substantially reduces the difficulty of singing pitch detection by eliminating a large number of wrong pitch candidates [1].

3. Literature review

There are various authentication methods which are summarized as follows:-

Chong Un ; Shih-Chien Yang, "A pitch extraction algorithm based on LPC inverse filtering and AMDF", IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. 25, No. 6, pp. 565 – 572, 1977. According to C. UN and S.C. Yang [17], four different types of methods namely, Autocorrelation based methods, linear predictive coding (LPC) – based methods, TD – based methods, Spectral analysis methods and data reduction methods [8].

A.L.C Wang. "Instantaneous and frequency-warped signal processing Technique for auditory source separation." Ph.D. dissertation, Dept. Elect. Eng., Stanford Univ., Stanford, CA, 1994. In this paper they developed a system for singing voice separation by using a Harmonic-locked loop technique to track a set of harmonicity related partials [7].

Y. Meron and K. Hirose, "Separation of singing and piano sounds," in Proc. 5th Int. Conf. Spoken Lang. Process. (ICSLP 98), 1998. The aim of this paper to separate singing voice from piano accompaniment. Also for this required a significant amount of prior knowledge is required, such as the partial tracks of premixing singing voice and piano or the music score for piano sound [6].

G. Hu and D. L. Wang, "Monaural speech segregation based on pitch tracking and amplitude Modulation," IEEE Trans. Neural Netw., vol. 15, no. 5, pp. 1135–1150, Sep. 2004.This system obtains its initial pitch estimation from the time lag corresponding to the maximum of a summary of autocorrelation function [5].

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. In this paper they propose Computational auditory scene analysis system which is effective for singing voice separation. In their system, pitch is detected based on a hidden markov model (HMM) [4].

Guoning Hu and Deliang Wang "A Tandem algorithm for pitch estimation and voice speech Segregation" IEEE Transaction on Audio, Speech, and Language Processing, VOL.18, NO.8 NOVEMBER 2010. In this they propose a new algorithm that achieves pitch estimation and speech segregation jointly and iteratively [3].

Morales-Cordovilla, J.A., Peinado, A.M.; Sanchez, V.; González, J.A., "Feature Extraction Based on Pitch-Synchronous Averaging for Robust Speech Recognition", IEEE Transactions on Audio, Speech, and Language Processing, Vol. 19, No. 3, pp. 640 – 651, March 2011. J. A. M. Cordovilla *et. al.* have worked on pitch – synchronous averaging to extract effectual features to precisely recognize speech. They have mainly focused on extracting the pitch from a noisy environment. Based on the pitch – synchronous averaging, two estimators have been proposed here [2].

Chao-Ling Hsu, DeLiang Wang, Jyh-Shing Roger Jang, and Ke Hu, "A Tandem Algorithm for Singing Pitch Extraction and Voice Separation from Music Accompaniment", IEEE Transactions on Audio, Speech, and Language Processing, Vol. 20, No. 5, p.p. 1482-1491, 2012. In this paper also tandem algorithm is used for pitch estimation and speech segregation but before this algorithm another one algorithm that is trend estimation is used to reduce the all the problem regarding the pitch estimation [1].

4. Proposed Work

4.1 Scope

A Singer voice extraction technique finds their wide scope in many speech analysis applications. The most popular application in which singer voice extraction found is used in Karoke. Karoke is not just a single application area to make use of singer voice recognition. It is a collection of various music processing systems to align lyrics to a singing voice. Few of such music processing systems include automatic lyrics identification, automatic singer identification, and many more. All these systems include pitch extraction processes and few may require singer voice recognition systems to make a successful karoke system. The proposed system extract voice signal from music as per the proposed diagram shown in figure1.

4.2Methodology

The methodology adopted is illustrated in Figure 1. It has four major components namely, Trend estimation, Mask estimation, pitch detection and supervised voice detection. Trend estimation helps to reduce false and artificial pitches arise due to music accompaniments or due to higher order harmonics. Subsequently, rough pitch ranges are estimated from the subjected speech signal by determining time – frequency (T – F) blocks. The second and third components constitute by using tandem algorithm.

In tandem algorithm, iterative processing is performed by considering the two processes in sequence, where the initial estimation is done through harmonic/percussive source separation (HPSS) process. The mask estimation includes estimating of Ideal Binary Mask (IBM), when the target pitch is applied. Based on the estimated IBM, pitch is determined. The recursive operation of Tandem algorithm for certain number of iterations enables detecting the pitch process.

The detected pitch is subjected to supervised voice extraction process. This work uses Support Vector Machine (SVM) to perform voice detection process. SVM is known for its outperformance, when it attempts to classify binary labels. Moreover, its supervisory nature further enhances the performance of voice detection performance. Being a supervised technique, SVM requires proper training to register its performance.

Hence, Figure 1 is portrayed with both offline and online process. The offline process represents the training phase for SVM, whereas the online process represents regular voice detection technique.

The proposed work concentrate on following points:-

1) Design of trend estimation for finding vocal component enhancement and singing pitch in a probable range.
2) Design of Tandem algorithm for Mask estimation and pitch detection.
3) Design of Support Vector Machine (SVM) for voice extraction.
4) Finally it is proposed to compare the results with existing system.

5. Proposed Block Diagram

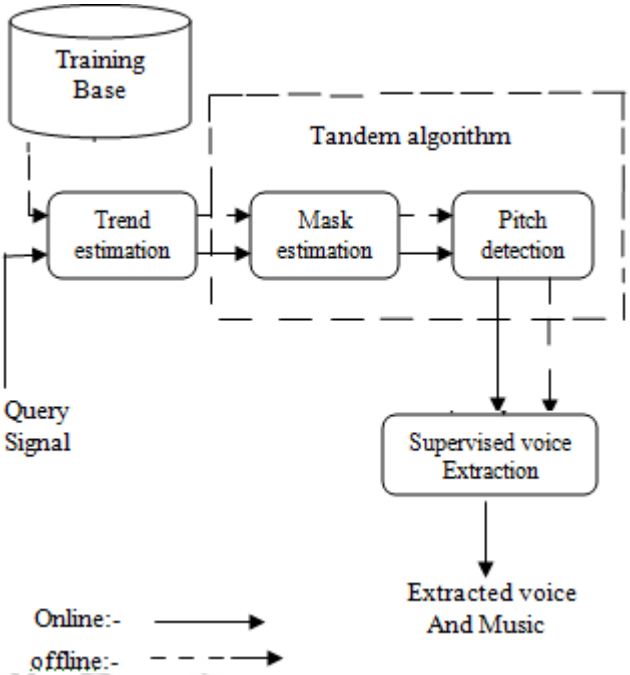

Figure 1: Overview of the proposed methodology for pitch estimation and singer voice Separation.

It is proposed to design whole system by using MATLAB software.

References

[1] Chao-Ling Hsu, DeLiang Wang, Jyh-Shing Roger Jang, and Ke Hu, "A Tandem Algorithm For Singing Pitch Extraction and Voice Separation from Music Accompaniment", IEEE Transactions on Audio, Speech, and Language Processing, Vol. 20, No. 5, p.p. 1482-1491, 2012.
[2] Morales-Cordovilla, J.A., Peinado, A.M. ; Sanchez, V. ; González, J.A., "Feature Extraction Based on Pitch-Synchronous Averaging for Robust Speech Recognition", IEEE Transactions On Audio, Speech, and Language Processing, Vol. 19, No. 3, pp. 640 – 651, March 2011.
[3] Guoning Hu and Deliang Wang "A Tandem algorithm for pitch estimation and voice speech Segregation" IEEE Transaction on Audio, Speech, and Language Processing, VOL.18, NO.8 NOVEMBER 2010.
[4] Yipeng Li, DeLiang Wang, Separation of Singing Voice from Music Accompaniment for Monaural Recordings, IEEE Transactions on Audio, Speech, and Language Processing, v.15 n.4, p.1475-1487, May 2007.
[5] G. Hu and D. L. Wang, "Monaural speech segregation based on pitch tracking and amplitude Modulation," IEEE Trans. Neural Netw., vol. 15, no. 5, pp. 1135–1150, Sep. 2004.
[6] Y. Meron and K. Hirose, "Separation of singing and Piano sounds," in Proc. 5th Int. Conf. Spoken Lang. Process. (ICSLP 98), 1998.
[7] A.L.C Wang. "Instantaneous and frequency-warped Signal processing Technique for auditory Source Separation." Ph.D. dissertation, Dept. Elect. Eng., Stanford Univ., Stanford, CA, 1994.

[8] Chong Un; Shih-Chien Yang, "A pitch extraction Algorithm based on LPC inverse filtering And AMDF", IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. 25, No.6, pp. 565 – 572, 1977.

Author Profile

Vikas Nichal received the BE degree in Electronics and tele-communication engineering from shivaji university, Kolhapur in 2012. He now is doing ME in Electronics and Tele-communication engineering from shivaji university, Kolhapur. Also working as assistant professor in Rajaram Shinde College of engineering, chiplun. His area of specialization in digital signal processing.

Mane V.A received the BE and ME in Electronics engineering. He is working as a assistant professor in Anasaheb Dange college of engineering, Ashta. His area of specialization in digital signal processing and Embedded system.

Gadhave D.D received the BE degree in Electronics engineering from shivaji university, Kolhapur in 2008. He now is doing ME in Electronics and Tele-communication engineering from Shivaji university, Kolhapur. Also working as assistant professor in Rajaram Shinde College of engineering, chiplun. His area of specialization in communication network and wireless network.