International Journal of Advanced Computer Technology (IJACT)
ISSN:2319-7900

Singing Voice Separation Using Hybrid ICA and Wavelet Thresholding

Shamili P, P G scholar, Vimal Jyothi Engineering College; Ancy K Sunny, Assistant Professor, Dept. of Computer Science, Vimal Jyothi Engineering College; T M Thasleema, Assistant Professor, Dept. of Computer Science, Central University of Kerala

Abstract

Separating singing voice from a mixture of singing voice and background music is an interesting problem in the domain of digital signal processing. There are many methods for separating singing voice from music. But separated singing voice contains background music in small amount. Combination of Independent Component Analysis (ICA) and wavelet thresholding can produce singing voice without background music. ICA takes mixtures of singing voice and background music and produces separated voice and music signal. Then wavelet thresholding method compute threshold from separated music signal and using that threshold, process separated singing voice. A comparison of five wavelets for this task is also conducted. Experiments show that haar wavelet provides good result.

Introduction

Singing voice separation from background music is a challenging problem. It belongs to blind source separation [1]. Blind source separation is the separation of a set of signals from set of mixed signals. The separation of signals is done without knowing about source signals. Separating singing voices from music accompaniment is required for many applications such as music information retrieval (MIR) studies like automatic lyrics recognition, identification of language of the song, automatic singer identification etc., and automatic karaoke generator. Singing voice separation is an interesting problem. Because there are some technical difficulties that make the singing voice separation is an interesting one. One of the difficulties is the similarity between singing voice and accompaniments, e.g., a piano, a guitar, and percussions. That is both the spectra of singing voice and harmonic instruments, such as a piano and a guitar, have harmonic structure. Therefore it is difficult for a simple harmonics extraction technique to detect only the singing voice in music signals. Second difficulty is that accompanying instruments do not satisfy some properties of "noise" such as whiteness and stationarity. Since music signals are not white noise and nor stationary, classical noise suppression technique will not work effectively in singing voice separation.

For singing voice separation, speech separation technique cannot be effective. Because the difference between singing voice and speech are significant. The main difference is the presence of an additional formant, known as the singing formant. This singing formant helps the voice of a singer to stand out from the accompaniment. However, the singing formant does not exist in other types of singing such as the ones in rock and country music. Another difference is based on the way of singing and speech is uttered. During singing, a singer often intentionally stretches the voiced sound and shrinks the unvoiced sound to match other musical instruments. This has two results. First, it changes the percentage of voiced and unvoiced sounds in singing. The majority of sounds generated during singing is voiced (about 90%), while speech has a larger amount of unvoiced sounds. Second, the pitch dynamics (the evolution of pitch in time) of singing voice tends to be piece-wise constant with abrupt pitch changes in between. This is opposite to the declination phenomenon in natural speech where pitch frequencies slowly drift down with smooth pitch change in an utterance. Besides these things, singing voice also has a wider pitch range. The pitch range of normal speech is between 80 and 400 Hz while the upper pitch range of singing can be as high as 1400 Hz. From the sound separation point of view, the major difference between singing and speech is the nature of other concurrent sounds. In a real acoustic environment, speech is usually altered by interference that can be harmonic or non-harmonic, narrowband or broadband. In most cases the interference is independent of speech in the sense that the spectral contents of target speech and interference are uncorrelated. For recorded singing voice, however, it is almost always accompanied by musical instruments that in most cases are harmonic, broadband, and are associated with singing since they are composed to be a coherent whole with the singing voice. These differences make the singing voice separation from the accompaniments potentially more challenging.

The implementation of this blind source separation is done by hybrid independent component analysis (ICA) and wavelet thresholding. The mixture of sources can be obtained by using two microphones that are placed at different distance from sources. ICA [2][3] takes song recorded at two microphones as inputs and produce separated singing voice and background music. The separated singing voice contains background music at low level. To remove that noise like background music from separated singing voice wavelet thresholding [4] is used. Threshold is calculated from separated background music signal.

Independent Component Analysis

Independent Component Analysis (ICA) is a method for extracting individual signals from mixtures of signals. It is

based on the assumption that different physical processes generate unrelated signals. ICA can be understood in terms of the classic 'cocktail party' problem, which ICA solves in an ingenious manner. Consider a cocktail party where many people are talking at the same time. If a microphone is present then its output is a mixture of voices. When given such a mixture, ICA identifies those individual signal components of the mixture that are unrelated. Given that the only unrelated signal components within the signal mixture are the voices of different people, this is precisely what ICA finds.

Let s_1 and s_2 are sources for singing voice and background music respectively. Two microphones are placed at a different distance from the sources. Then songs recorded at two microphones are of different amplitude. Let x_1 and x_2 are songs captured by the two microphones. The relative amplitude of each source at the microphone is related to the microphone-source distance and can be defined as a weighing vector A_{ij} for each source.

$$x=As$$

Where $x=(x_1, x_2)$, $s= (s_1, s_2)$ and A is the mixing matrix that specifies the relative contributions of the source signals s to each mixture x_i. The matrix A defines a linear transformation on the signals s. Such linear transformations can usually be reversed in order to obtain an estimate u of source signals s from signal mixtures x.

$$u=Wx$$

Where $u= (u_1, u_2)$ and W is the separating matrix and it is the inverse of A. Since the mixing matrix A is not known, it cannot be used to find W. The solution to this problem is adjust W so as to make the estimated source signals u mutually independent. This is achieved by adjusting W to maximize the joint entropy probability density function of u.

Wavelet Thresholding

Wavelet thresholding is a denoising technique [5] that exploits the features of wavelet transform [6][7]. It removes the noise by eliminating coefficient below some threshold value. Here threshold is calculated from background music obtained ICA. Wavelet thresholding consist of following steps

1. Perform wavelet packet decomposition [8] on background music obtained from ICA
2. Calculate threshold value using threshold selection methods [9].
3. Perform wavelet packet decomposition on singing voice obtained from ICA.
4. Apply thresholding function on coefficients of singing voice
5. Perform wavelet packet reconstruction to get clean audible singing voice

Simulation Results

Simulation results have been carried on 30 songs that are extracted from MIR-1 K database [10]. MIR-1 K database consists of 1000 songs of 16 kHz sampling frequency. Each song consists of background music and singing voice recorded at left and right channels respectively. For each song background music and singing voice are mixed using a constant mixing matrix so as to generate a pair of mixtures. Soft thresholding with universal threshold is used. Wavelet thresholding is done using five wavelets namely haar, db4, sym4, coif1 and bior 6.8. Results are shown through comparison among them. The parameter used for comparison is Signal to Noise Ratio (SNR).

$$SNR=10 \log_{10}(var(s_1)/var(s_1-u_1))$$

Where u_1 is the independent component corresponding to singing voice and x is the mixture.

1 SongNo	2 ICA	3 haar	4 db4	5 sym4	6 coif1	7 bior68
1	-5.4589	-4.3934	-4.8607	-4.8483	-4.8031	-4.8926
2	-5.7630	-4.9811	-5.2196	-5.2183	-5.1748	-5.2625
3	-5.3863	-4.5854	-4.9552	-4.9627	-4.8954	-4.9997
4	-5.4511	-4.6564	-4.9878	-4.9836	-4.9248	-5.0214
5	-5.5557	-4.8242	-5.1445	-5.1415	-5.0919	-5.1773
6	-6.8920	-6.5288	-6.8271	-6.8267	-6.7871	-6.8399
7	-5.9497	-5.4127	-5.8764	-5.8758	-5.8273	-5.8883
8	-5.7721	-4.9525	-5.1388	-5.1320	-5.1119	-5.1551
9	-5.7915	-5.0616	-5.2354	-5.2368	-5.2119	-5.2806
10	-6.1082	-5.5450	-5.6868	-5.6762	-5.6535	-5.7105

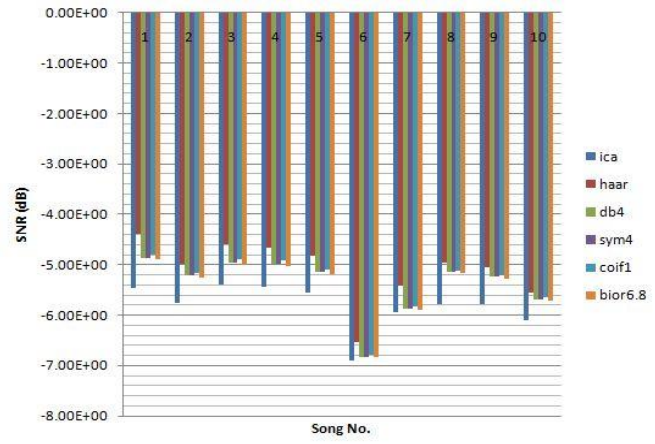

Figure 1: Comparison of SNR for ICA and hybrid ICA and wavelet thresholding using different wavelets at decomposition level1 for 10 songs.

Figure 1 shows the comparison of SNR for ICA and hybrid ICA and wavelet thresholding using different wavelets at decomposition level 1.

It shows that hybrid ICA and wavelet thresholding performs better than ICA. The proposed method achieves great-

ISSN:2319-7900

er SNR than ICA. And this SNR varies with type of wavelet. From figure it is clear that haar wavelet provide greater SNR than other wavelets. So haar wavelet is suitable for this task.

Conclusion

Hybrid ICA and wavelet thresholding can separate singing voice from background music. The results conclude that it can perform better than ICA. Wavelet thresholding is done using five wavelets namely haar, db4, sym4, coif1and bior 6.8. Out of which hybrid ICA and wavelet thresholding with haar wavelet provides better SNR than others.

References

[1] A. Mary, Anto Prem Kumar, Anish Abraham Chacko, "Blind Source Separation Using Wavelets", IEEE International Conference on Computational Intelligence and Computing Research, ISBN 97881 8371 3627, 2010
[2] James V Stone, "Independent Component Analysis: an introduction", Trends in Cognitive Sciences, Vol. 6, No. 2, February 2002.
[3] K Prakash, Hepzibha Rani D, "Blind Source Separation for Speech Music and Speech Speech Mixtures", International Journal of Computer Applications, Vol. 10, No. 12, January 2015
[4] Jeena Joy, Salice Peter, Neetha John, "Denoising using soft thresholding", International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, Vol. 2, Issue 3, March 2013
[5] David L Donoho, "Denoising by Soft Thresholding", IEEE Transactions on Information Theory, Vol. 41, No. 3, May 1995
[6] Christopher E Heil, David F Walnut, "Continuous and Discrete Wavelet Transform", Society for Industrial and Applied Mathematics, Vol. 31, No. 4, December 1989
[7]https://en.wikipedia.org/wiki/Wavelet_transform
[8] Monika Sheron, Sanjeev Kumar, Amod Kumar, "Wavelet ICA Based Denoising of Electroencephalogram", International Journal of Information and Computation Technology, Vol. 4, No. 12, 2014, pp 1205-1210
[9] Guomin, Daming Zhang, "Wavelet Denoising" Nanyang echnological University, Singapore
[10]https://sites.google.com/site/unvoicedsoundseparation/mir-1k