International Journal of Scientific Engineering and Technology Research

ISSN 2319-8885
Vol.03,Issue.30
October-2014,
Pages:5933-5934

www.ijsetr.com

Effect of Similarity Measures on Repetitive Audio Source Separation

MITA SURESH[1], INDU REENA VARUGHESE[2]

[1]PG Scholar, Amal Jyothi College of Engineering, Kottayam, Kerala, India, Email: mitasuresh@gmail.com.
[2]Assistant Professor, Dept of ECE, Amal Jyothi College of Engineering, Kottayam, Kerala, India.

Abstract: Separating the vocal signal from background music is important for many applications, such as content-based search for singer identification and lyrics recognition. The technique of considering popular music as a superimposition of locally non-repeating voice signal on a repeating musical structure has recently put forth competitive single channel voice extraction methods. The basic idea of these algorithms is to find the most similar k-frames to every time-frequency frame, and use the optimized version of these to model as the background music. A time-frequency mask is then modeled from the repeating background channel to filter out the sources. This paper analyzes the effect distance metrics have on the separation performance when different similarity matrices are used for modeling the repeating structures.

Keywords: Blind Audio Source Separation; Distance Metric; Similarity Matrix.

I. INTRODUCTION

The removal of vocals from a musical recording is traditionally accomplished in stereophonic tracks wherein the vocals are exactly the same in both stereo channels (the vocals are mixed at exactly the center) by subtracting one channel over the other. Often, the lower end of the spectrum is also removed along with the vocals and this resulted in a distorted output signal. There are different needs for setting about this task and once the different sources are separated, the individual sources can be used as the input to several other tasks. Primarily, it is the melody which is extracted for various purposes such as elimination of the background noise (in which case, not much consideration is given to ensuring the quality of the other sources in the mixture) in hearing aids, segmenting the audio for music summarization, content-based music indexing or searching (transcription, instrument identification, singer identification, lyrics transcription), object-based coding, robust speech recognition and other audio manipulations. Audio source separation is often used as pre-processing or post-processing step. Music source separation is relevant to many applications of digital signal processing.

II. PREVIOUS WORK

This section briefly discusses various methods by which blind audio source separation (BASS) has been achieved. Though source separation has been discussed for decades and several methods have been proposed independently from researchers from different communities, there is no systematic summarization which focuses on music source separation. There are mainly three families of methods in blind signal separation: Independent Component Analysis (ICA), in which sources are statistically independent, stationary and at most one of them is Gaussian (in their basic versions); Sparse Component Analysis (SCA), in which the sources are assumed to be sparse i.e. most of the samples are null or close to zero; Non-negative Matrix Factorization (NMF): in which both the sources in the mixture are positive, with possible sparse constraints. A more conclusive survey on this expanding research field can be found in the referenced review papers [3]. A novel way[3] of extracting the main vocal signal from the accompaniment is to extract the non-repeating structures in the audio. Theoretically, it is repetition that develops the smallest element of music called motive. Moreover, repetition has also been used to reveal the syntax in music signals. Advantages offered by this time-frequency domain technique include non-requirement of information regarding the input signals and simplicity of the algorithm due to any complex probabilistic framework.

The work has been extended in [2] to accommodate music who's repeating background score change with time. Further, the online REPET [1] examines the case when the background has repeating structures occur intermittently, as opposed to a global or local periodicity. The technique is simple enough to offer real-time computation by buffering time-frequency frames of the input in time. Each frame in the spectrogram is compred to the buffered frames using a distance metric and sorted according to similarity. The work done in this paper is to compare the quality of the separated output when the distance metric is varied. The relevance of the study is validated by the superior results obtained in [1], which blindly used cosine distance.

III. EXPERIMENTAL SETUP

The performance of the algorithm has been qualitatively evaluated using listening tests and quantitatively measured using the BSS_EVAL toolbox [6] for two song clips of the

MIR-1K dataset [5], featuring one male vocal track (`abjones_1_01.wav') and the other, one female vocal track (`amy_1_0.wav') for the four standard distance measures-cosine, Manhattan, Euclidean and Chebyshev[4]. The toolbox defines four signal quality parameters: Signal to Distortion Ratio (SDR), Signal to Interference Ratio (SIR) and Signal to Artefact Ratio (SAR) as defined in [7].

IV. RESULTS OF EXPERIMENT

The results have been tabulated in Tables I to IV with input audio clips mixed linearly and instantaneously into a monaural mixture using three different `voice-to-music' ratios: -5dB (music is louder), 0 dB (same original level), and 5 dB (voice is louder). The results of the performance measurement show that as a general trend, using the Chebyshev distance for similarity indication yielded the highest SDR value for both vocal and background track when the vocal track energy was not predominated by the background, i.e. when the mixing ratio was 0 dB and -5dB. When the MR is 5 dB (music is much louder than vocals), cosine distance shows the highest value of SDR. Also, the SIR value of the estimated vocal track is highest when the Euclidean distance is used. The results hold true in all tested samples of the MIR-1K database, providing sufficient evidence to indicate the superior performance in using Chebyshev distance metric for audio separation.

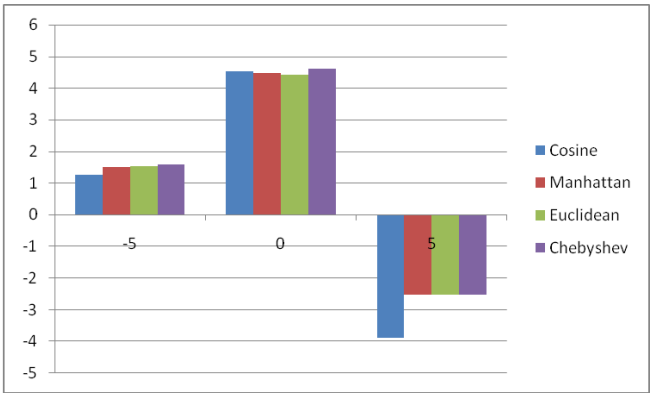

Figure1.Performance evaluation of estimated and original background tracks expressed in SDR (x-axis) over mixing ratios of -5dB, 0dB and 5dB.

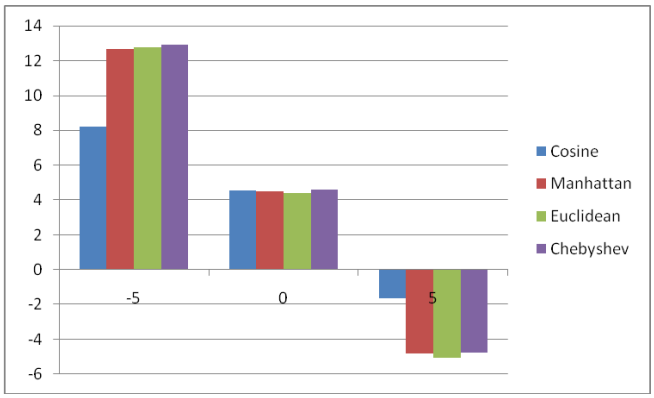

Figure2.Performance evaluation of estimated and original vocal tracks expressed in SDR (x-axis) over mixing ratios of -5dB, 0dB and 5dB.

Table1: SDR of the Background Track Evaluated over Different Distance Metrics Averaged over The Input Audio Clips.

Mixing Ratio(dB)	Cosine	Manhattan	Euclidean	Chebyshev
-5	1.2702	1.51625	1.523	1.58885
0	4.5293	4.48795	4.4096	4.6024
5	-3.891	-2.5127	-2.5131	-2.5128

Table2: SDR of The Vocal Tracks Evaluated over Different Distance Metrics Averaged over The Input Audio Clips.

Mixing Ratio (dB)	Cosine	Manhattan	Euclidean	Chebyshev
-5	8.19495	12.6787	12.7711	12.9125
0	4.52925	4.48795	4.4096	4.6024
5	-1.6238	-4.7987	-5.0796	-4.7639

V. CONCLUSION

The quantitative results illustrate how the performance of the online REPET algorithm differs when similarity is measured by using cosine, Manhattan, Euclidean and Chebyshev distances. Employing the Chebyshev distance had resulted in higher value of the SDR ratio, when averaged over all test inputs of the MIR-1K database, showing improved separation in the case where vocal energy is not predominated by accompaniment track.

VI. REFERENCES

[1] Z. Rafii and B. Pardo, “Online REPET-SIM for real-time speech enhancement,” 38th International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, May 26-31, 2013.

[2] Z. Rafii, and B. Pardo, “Music/voice separation using the similarity matrix,” 13th International Society for Music Information Retrieval, Porto, Portugal, October 8-12, 2012.

[3] Z. Rafii and B. Pardo, “REpeating Pattern Extraction Technique (REPET): A simple method for music/voice separation,” IEEE Transactions on Audio, Speech, and Language Processing}, Volume 21, Issue 1, pp. 71-82, January, 2013.

[4] H. Cha, “Comprehensive survey on distance/dissimilarity measures between probability density functions,” International Journal of Mathematical Models and Methods in Applied Sciences, no.4, 2007.

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

[6] C. Fevotte, R. Gribonval, and E. Vincent, “BSS_EVAL Toolbox User Guide", IRISA, Rennes, France, 2005, Tech. Rep. 1706.

[7] E. Vincent, R. Gribonval, C. Fevotte, “Performance measurement in blind audio source separation,” IEEE Trans.