Musical Instrument Signal Separation from Popular Songs

Seung Hoon Choi
Fall 2009 CS 229 Final Project
Final Report

December 11, 2009

1 Introduction

Signal separation from mix has been an interest of research in recent years, and many learning algorithms and variations have been developed. In this paper we focus on separation of specific type of signal with practical application: musical instrument signal separation. The result of such separation can be used by amateur instrument players to listen to their part of a given music more clearly, and learn how to play their part. We experiment with existing learning algorithms to separate a single musical instrument from popular music MP3 file. This experiment will hopefully reveal particular nature of the target instrument and we will be able to extend and modify the learning algorithm to get better performance.

2 Experimentation

Idea is to use blind source separation unsupervised learning algorithms. For this idea, we start with independent component analysis, and follow with different and/or extended algorithms to get better performance. Many variations of ICA algorithms are implemented in ICALAB software package [2][1]. ICALAB is used extensively for the experiment.

3 Using standard ICA

For the experiment, we use MP3 file of Nirvana's Smells Like Teen Spirit. The following two channels are first several seconds of the song:

Using off-the-shelf ICA algorithm returns two signals below:

Though it is hard to see from the signal outputs, the first signal is almost identcal to the original music, while the second result sounds louder and distorted when confirmed by converting back to mp3 and play it.

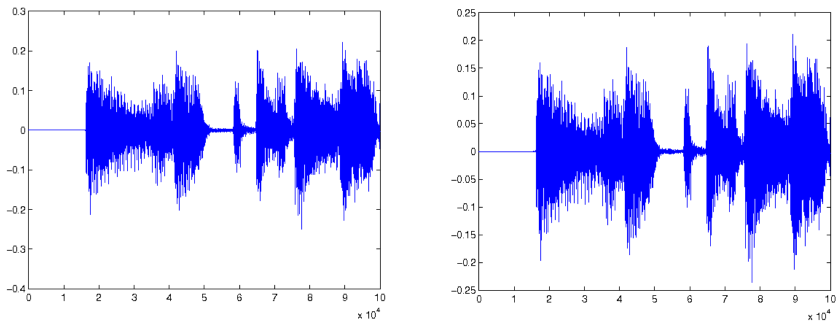

Figure 1: Channel 1 (left) and channel 2 (right) of MP3

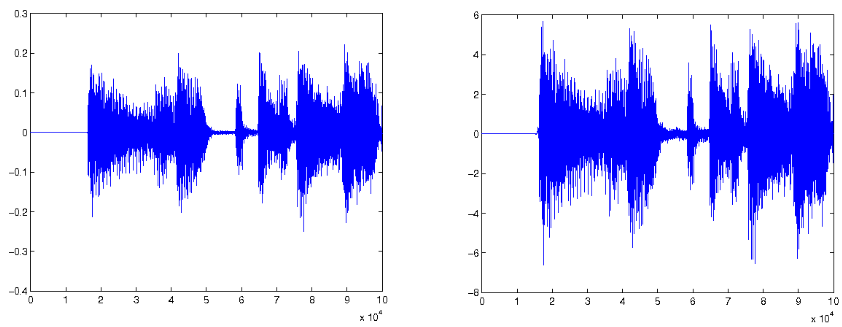

Figure 2: Result

4 Dynamical Embedding

An interesting application of ICA has been studied in neuroscience area in [3]. In the paper the authors apply standard ICA algorithm to separate out different brain signals from a single mixture of brain signals. They do so by using a technique called dynamical embedding. Assumption of dynamic embedding is that the measured signal is due to non-linear interaction of just a few degrees of freedome, with additive noise. This suggests the existence of an unobservable deterministic generator of the observed data. This is applied to music signal to separate out independent components.

Basic idea is to construct dynamic embedding matrix out of a series of delay vectors from the mixture, then feed it to standard ICA algorithm. One way to construct a DE matrix is from a series of delay vectors taken from the observed data x(t), where the state of the unobservable system at time t, X(t), is given by

$$X(t) = x(t-\tau), x(t-2\tau), \cdots, x(t-(m-1)\tau) \tag{1}$$

where τ is the lag and m is the number of lags or the embedding dimension. Let N stand for the number of delay vectors then DE matrix is given by

$$X = \begin{bmatrix} x_t & x_{t+\tau} & \cdots & x_{t+N\tau} \\ x_{t+\tau} & x_{t+2\tau} & \cdots & x_{t+(N+1)\tau} \\ \vdots & \vdots & \ddots & \vdots \\ x_{t+(m-1)\tau} & x_{t+m\tau} & \cdots & x_{t+(m+N-1)\tau} \end{bmatrix} \tag{2}$$

Each delay vectors of the observed signal is pictorially:

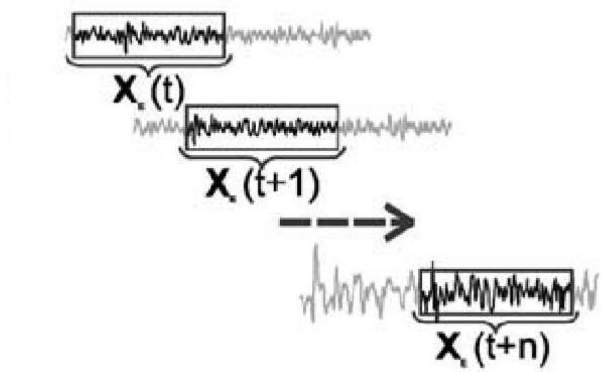

Figure 3: Delay vectors from observed signal

5 Results

In order to ease the inspection of results in signals, I have instead run the algorithm with a mixture of two signals, one signal representing a vocal sound, the other representing musical instrument playing along with the vocal. Also, these are truly single-channel wave file format. The following represents the two original signals, and mixture of the two:

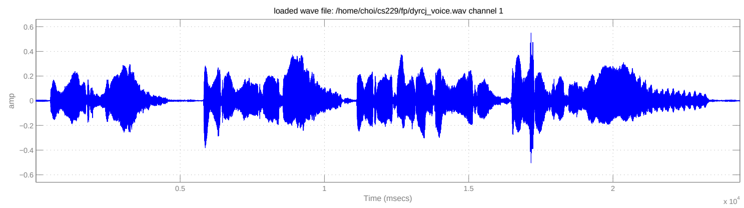

Figure 4: Original vocal signal

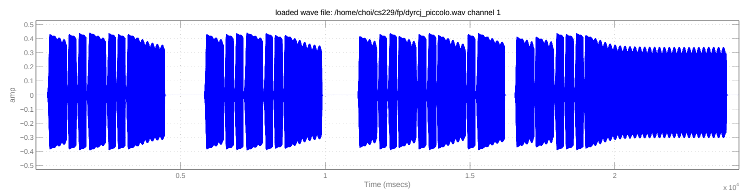

Figure 5: Original musical instrument signal

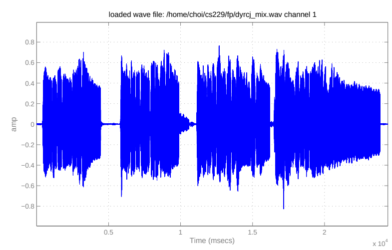

Figure 6: Mixture of the two signals

The delay vectors, and DE matrix is generated with $\tau = 1$ and $m = 2$ (two original signals), and using ICA algorithm, the result is following:

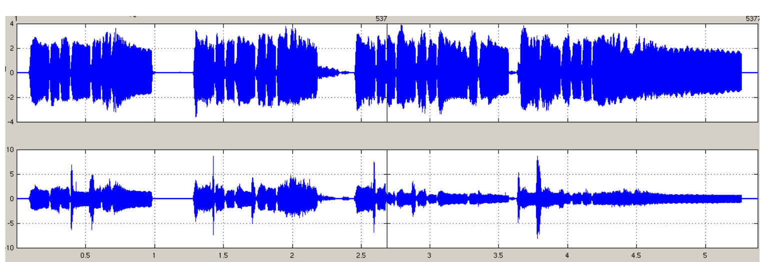

Figure 7: Results

6 Conclusion

The result signals show some resemblance to the original signals, and it is interesting to use the standard ICA algorithm, which usually requires as many signals as the number of sources, to separate a single-channel mixture without changes to the algorithm. Although the results are not as clean as some other more complex techniques, it should be interesting to see how changing the parameters could give better results.

References

[1] A. Cichocki and S. Amari. Adaptive blind signal and image processing: Learning algorithms and applications, 2003.

[2] A. Cichocki, S. Amari, K. Siwek, T. Tanaka, and A. H. P. et al. Icalab toolboxes. http://www.bsp.brain.riken.jp/ICALAB.

[3] C. J. James and D. Lowe. Extracting multisource brain activity from a single electromagnetic channel. *Artificial Intelligence in Medicine*, 2003.