

# An Efficient Activation Function for Blind Separation of Speech Signals

Shouyu Sun

Beijing Information Technology Institute, 100094, China

## Abstract

The aim of this paper is to study blind separation of speech signals based on feedforward neural network with sine activation function. The natural gradient is a major contribution to blind source separation (BSS) because its performance is independent of the mixing matrix. Computer simulations are provided by speech signals as sources (ten speakers) and in the instantaneous mixing case show that the algorithm with sine activation function converges faster and recovers independent sources better than other activation functions.

**Keywords:** Blind Separation, Speech Signals, Activation Function.

## 1. Introduction

Since the pioneering work of Jutten and Herault [1], the problem of separating out linear instantaneously mixed statistically independent source signals by the independent component analysis (ICA) technique has been widely investigated both in the neural network community, and in the signal processing community and several algorithms and methods have been proposed by many authors. An excellent information-theoretic approach by Bell and Sejnowski [2] with the natural gradient modification developed by Amari [3] has attracted much interest in the neural network community.

The aim of this paper is to study blind separation of speech signals based on

feedforward neural network with sine activation function. Compared to the hyperbolic tangent activation function, sine activation function makes the algorithm convergence faster.

## 2. ICA Technique

The task of source separation is to estimate the mixing matrix or to recover original sources, given only their mixtures.

For separating out  $N$  linearly mixed independent sources, we use a fully connected feedforward neural network with  $N$  inputs and  $N$  outputs shown in Fig.1 whose linear part is described by the relationship:

$$\mathbf{u}(n) = \mathbf{W}\mathbf{x}(n) \quad (1)$$

where  $\mathbf{x}(n) = [x_1(n) \ x_2(n) \cdots x_N(n)]^T$  is the network input vector,

$\mathbf{u}(n) = [u_1(n) \ u_2(n) \cdots u_N(n)]^T$  denotes the network output vector, and  $\mathbf{W}$  is the weight-matrix. Each row-vector

$\mathbf{w}_j$  represents the weight-vector of corresponding  $j$ -th neuron. The mixing model is

$$\mathbf{x}(n) = \mathbf{A}\mathbf{s}(n) \quad (2)$$

where  $\mathbf{A}$  is a constant real-valued full-rank  $N \times N$  unknown mixing matrix,  $\mathbf{s}(n) = [s_1(n) \ s_2(n) \cdots s_N(n)]^T$  is a  $N \times 1$  vector containing  $N$  source signals to be separated, i.e. we only deal

with the instantaneous linear mixture problem.

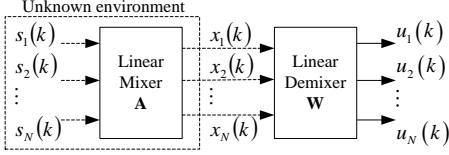


Fig.1: Block diagram illustrating basic BSS problem.

The basic principle that the ICA technique is based on is that after application of mixing model (2) the resulting processes  $x_j$  are no longer statistically independent. In order to achieve separation, that means finding a weight  $\mathbf{W}$  such that  $\mathbf{WA}$  equals the identity matrix, the weight matrix may be learnt so that the network's outputs (1) become as independent as possible.

The natural gradient algorithm was first proposed by Cichcki [5] and later studied and justified by Amari [3]. The natural gradient algorithm is

$$\begin{aligned} \mathbf{W}(n+1) &= \mathbf{W}(n) \\ &+ \mu \{ \mathbf{I} - \Phi[\mathbf{u}(n)] \mathbf{u}^T(n) \} \mathbf{W}(n) \end{aligned} \quad (3)$$

Where  $\Phi(\mathbf{u})$  is also called as the activation function of the learning algorithm [4].

### 3. Simulations

We apply the natural gradient algorithm Eq. (3) to separate sources with different activation function to compare their performance. A typical activation function is sigmoid function used widely in many papers

$$\Phi(\mathbf{u}) = 2 \tanh(\mathbf{u}) \quad (4)$$

The sine activation function which is defined as

$$\Phi(\mathbf{u}) = \begin{cases} 1 & \mathbf{u} > \pi/4 \\ \sin(2\mathbf{u}) & -\pi/4 \leq \mathbf{u} \leq \pi/4 \\ -1 & \mathbf{u} < -\pi/4 \end{cases} \quad (5)$$

is used in this paper.

In this simulation we use natural speech as our sources. We use ten different speakers on 4 seconds speech segments, which were sampled at 8kHz and were normalized such that their variance is equal to 1. A nonsingular random mixing matrix  $\mathbf{A}$  was generated with values uniformly distributed between  $-1$  and  $1$  to make the mixed time series  $\mathbf{x}$  from the original sources  $\mathbf{s}$ . To measure the performance of the natural gradient algorithm with two different activation function (4) and (5), we use the cross-talking error [7]

$$E = \sum_{i=1}^N \left( \sum_{j=1}^N \frac{|p_{ij}|}{\max_k |p_{ik}|} - 1 \right) + \sum_{j=1}^N \left( \sum_{i=1}^N \frac{|p_{ij}|}{\max_k |p_{kj}|} - 1 \right) \quad (6)$$

where  $\mathbf{P} = (\mathbf{p}_{ij}) = \mathbf{WA}$ .

The parameter settings chosen for simulation of Fig.2 are  $\mu = 0.0002$ ,  $\mathbf{W}(0) = 0.1\mathbf{I}$ , number of passes is 3 ( $3 \times 4 \text{ seconds}$ ). We can see from Fig.2 that the algorithm with sine activation function outperforms the algorithm with sigmoid function. We could have taken larger adaptation steps to speed up the convergence but the separation would then not have been that good.

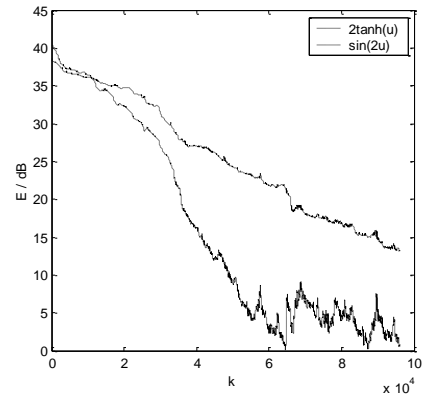
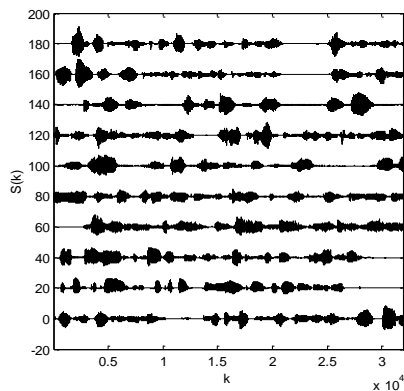


Fig.2: Comparison of performance using activation function  $\varphi(u) = 2 \tanh(2u)$  and  $\varphi(u) = \sin(2u)$

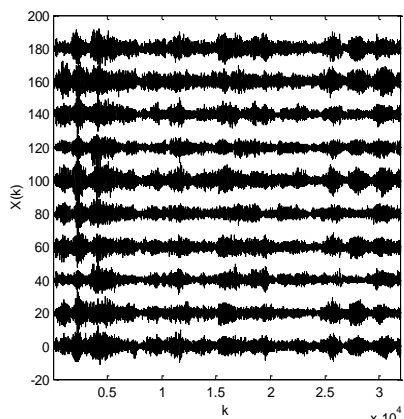
After the algorithm's convergence, performance matrix  $\mathbf{P}$  is

0.0048	-0.0065	0.0027	0.0105	0.0125	<u>-2.1222</u>	-0.0072	-0.0037	0.0259	0.0234
0.0003	-0.0171	0.0324	0.0027	-0.0130	0.0015	0.0007	<u>-2.2270</u>	-0.0055	-0.0045
-0.0017	-0.0233	-0.0260	0.0073	-0.0432	-0.0006	<u>-1.6746</u>	0.0083	-0.0261	-0.0134
-0.0083	-0.0212	-0.0403	-0.0008	<u>1.7027</u>	-0.0232	-0.0420	-0.0035	-0.0060	-0.0423
0.0368	0.0121	0.0029	<u>1.4724</u>	0.0255	-0.0091	0.0027	0.0041	-0.0072	0.0295
0.0007	<u>4.3026</u>	0.0112	-0.0008	0.0031	0.0007	-0.0015	0.0005	0.0011	0.0032
0.0000	-0.0102	<u>-4.1460</u>	0.0005	-0.0003	0.0006	-0.0023	-0.0005	-0.0007	-0.0118
0.0031	-0.0077	0.0258	-0.0098	0.0105	0.0260	-0.0089	-0.0037	<u>1.7354</u>	0.0227
<u>-1.2399</u>	0.0209	0.0029	-0.0319	0.0060	0.0114	0.0037	0.0131	<u>0.0064</u>	-0.0134
-0.0091	-0.0021	0.0043	0.0029	-0.0233	0.0101	0.0060	-0.0109	-0.0337	<u>-1.8943</u>

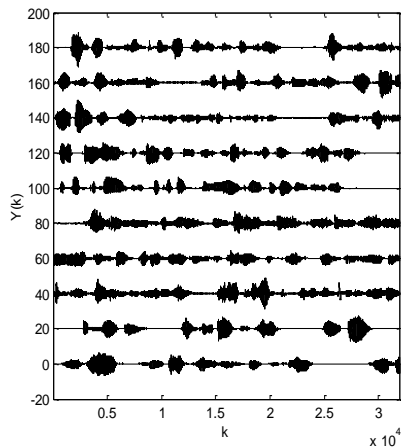
As can be seen, only one substantial entry (underlined) exists in each row and column. Fig.3 shows the simulation result of ten speakers for blind source separation with  $\varphi(u) = \sin(2u)$ .



a. Original sources



b. Mixed signals



c. Recovered sources

Fig.3: the example of ten speakers for blind source separation with  $\varphi(u) = \sin(2u)$ .

#### 4. Conclusions

We have presented an efficient activation function for blind source separation based on fully connected feedforward network, when sources are nonstationary. Through computer simulations, we have demonstrated that the natural gradient algorithm with sine activation function converges faster and recovers independent sources better than other functions.

#### 5. References

- [1] Jutten, C., Herault, J., "Blind separation of sources, part I: an adaptive algorithm based on neuromimetic architecture," *Signal processing*, vol. 24, pp. 1-10, 1991.
- [2] A. J. Bell and T. J. Sejnowski, "An information-maximization approach for blind separation and blind deconvolution," *Neural Computation*, vol. 7(6), pp. 1126-1159, 1995.
- [3] Yang, H. H., Amari, S.-I., "Adaptive online learning algorithms for blind separation: maximum entropy and minimum mutual information," *Neu-*

- ral Computation*, vol. 9, pp. 1457-1482, 1997.
- [4] Simon Haykin, *Neural Networks: a comprehensive foundation*, 2<sup>nd</sup> ed, Tsinghua University Press, pp.511-541, 1999.
  - [5] A. Cichocki, R. Unbehauen, L. Moszczynski, and E. Rummert, "A new online adaptive learning algorithm for blind separation of source signals," *Proc. ISANN*, Taiwan, pp. 406-411, 1994.
  - [6] S. Fiori, "Blind signal processing by the adaptive activation function neurons," *Neural Networks*, 13, pp. 597-611, 2000.
  - [7] Jacob Benesty, "An introduction to blind source separation of speech signals," *Acoustic signal processing for telecommunication*, edited by Steven L. Gay, Jacob Benesty, Boston: Kluwer Academic, pp. 321-329, 2000.
  - [8] Seungjin Choi, Andrzej Cichocki, Shunichi Amari, "Equivariant nonstationary source separation," *Neural Networks*, 15, pp. 121-130, 2002.
  - [9] Kuo-Kai Shyu, Ming-Huan Lee, Yu-Te Wu, Po-Lei Lee, "Implementation of pipelined FastICA on FPGA for Real-time Blind source separation," *IEEE Trans. Neural Netw.*, vol.19, no.6, June 2008.
  - [10] J.T. Chien and B.C. Chen, "A new independent component analysis for speech recognition and separation," *IEEE Trans. Speech Audio Process.*, vol.14, no.4, pp. 1245-1254, July 2006.