A POLAR-COORDINATE BASED ACTIVATION FUNCTION FOR FREQUENCY DOMAIN BLIND SOURCE SEPARATION

Hiroshi Sawada *Ryo Mukai* *Shoko Araki* *Shoji Makino*

NTT Communication Science Laboratories
2-4 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0237, Japan
{sawada,ryo,shoko,maki}@cslab.kecl.ntt.co.jp

ABSTRACT

This paper presents a new activation function for an ICA algorithm to process complex-valued signals, which is used in frequency domain blind source separation. The new activation function is based on the polar coordinates of a complex number, whereas the conventional one is based on the Cartesian coordinates of a complex number and calculates the real part and imaginary part separately. The new activation function eliminates an undesirable constraint occurred by the conventional function. In experiments for separating speech signals in a reverberant environment, we obtained improved SNRs by using the new activation function.

1. INTRODUCTION

Blind source separation (BSS) is a technique to estimate original source signals using only sensor observations that are linear mixtures of the original signals. Independent component analysis (ICA) [1–4] works very well for BSS, if the mixture is instantaneous (non-convolutive). In a real room environment, however, sounds are mixed in a convolutive manner with reverberation, and longer reverberation makes a BSS problem more difficult. Several methods have been proposed for convolutive mixtures. One of the major methods is frequency domain BSS [5–14].

In frequency domain BSS, a convolutive mixture problem of the time domain is converted into multiple instantaneous mixture problems of the frequency domain. Then, these instantaneous mixture problems are solved by ICA in every frequency bin. Since the conversion is performed by using a windowed discrete Fourier transform (DFT), we have to deal with complex numbers. An activation function for an ICA algorithm to process complex numbers was proposed [5]. This function, however, imposes an additional constraint that prevents a learning algorithm from converging. To avoid the additional constraint, another formula to calculate a gradient of an ICA algorithm was proposed [6, 7], and has been used in [8, 9, 13, 14].

In this paper, we propose a new activation function for an ICA algorithm to process complex numbers. It is based on the polar coordinates of a complex number, and eliminates the additional constraint discussed above. Experimental results have shown that the new activation function works well and SNRs (signal-to-noise ratios) are improved over conventional methods. We discuss several reasons why improved results were obtained, by looking into what happens in ICA algorithms.

In Section 2, we explain frequency domain BSS in detail. The new activation function is proposed in Section 3. In Section 4, we show and discuss experimental results. We conclude this paper in Section 5.

2. FREQUENCY DOMAIN BSS

2.1. Problem Formulation

Suppose that there are N source signals $s_p(t),\ (1 \leq p \leq N)$ that are mutually independent, and these signals are observed at M microphones in a real room environment with reverberation. The observed signals cab be written in a convolutive mixture form:

$$x_q(t) = \sum_{p=1}^{N} \boldsymbol{h}_{qp} * s_p(t),\ (1 \leq q \leq M),$$

where $\boldsymbol{h}_{qp}$ represents the impulse response from source p to microphone q, and $*$ denotes the convolution operator.

The goal of blind source separation is to separate observed signals $x_q(t)$ into N unmixed signals $y_p(t),\ (1 \leq p \leq N)$ that are mutually independent. The separation has to be done without knowing impulse responses $\boldsymbol{h}_{qp}$ nor original source signals $s_p(t)$. The unmixing system can be composed of $N \times M$ FIR filters. The unmixed signals are obtained by

$$y_p(t) = \sum_{q=1}^{M} \boldsymbol{w}_{pq} * x_q(t),\ (1 \leq p \leq N),$$

where $\boldsymbol{w}_{pq}$ represents the coefficients of the FIR filters. Figure 1 shows a BSS system for the case of $N = M = 2$.

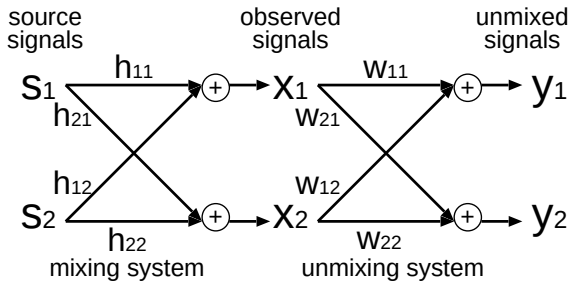

Fig. 1. Model of a BSS system

2.2. Framework of Frequency Domain BSS

A convolutive mixture in the time domain corresponds to an instantaneous mixture in the frequency domain. Therefore, we can apply an ordinary ICA algorithm in the frequency domain to solve a BSS problem. A T-point windowed DFT is used to convert time domain signals $x_q(t)$ into frequency domain time-series signals:

$$X_q(\omega, m) = \sum_{t=0}^{T-1} e^{-j\omega t}\, x_q(t)\, win(t - mS), \quad (\omega = 0, \frac{1}{T}2\pi, \ldots, \frac{T-1}{T}2\pi)$$

where win denotes a window function and S is the shifting interval of the window. The frame length T of the window is the same as the number of frequency bins, and also the length of the FIR filters $\boldsymbol{w}_{pq}$ composing the unmixing system.

For each frequency ω, an ICA algorithm is applied to obtain an unmixing $N \times M$ matrix $\boldsymbol{W}(\omega)$ and frequency domain N signals $\boldsymbol{Y}(\omega, m)$, which are estimations of the source signals at the frequency:

$$\boldsymbol{Y}(\omega, m) = \boldsymbol{W}(\omega)\boldsymbol{X}(\omega, m),$$

where $\boldsymbol{X}(\omega, m) = [X_1(\omega, m), \ldots, X_M(\omega, m)]^T$. Then, we can obtain FIR filters $\boldsymbol{w}_{pq}$ of length T by applying the inverse DFT to all $\boldsymbol{W}(\omega)$.

2.3. ICA algorithm

In an ICA algorithm, an unmixing matrix $\boldsymbol{W}$ is gradually improved by the learning rule:

$$\boldsymbol{W}_{i+1} = \boldsymbol{W}_i + \Delta\boldsymbol{W}_i$$

based on the minimization of the mutual information of $\boldsymbol{Y}$ [1, 2]. For the calculation of $\Delta\boldsymbol{W}$, the natural gradient [3] is widely used:

$$\Delta\boldsymbol{W} = \mu\,[I - \langle\Phi(\boldsymbol{Y})\boldsymbol{Y}^T\rangle]\,\boldsymbol{W}.$$

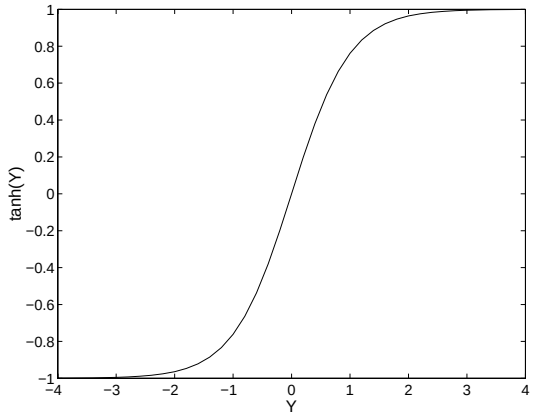

Fig. 2. Hyperbolic tangent $\tanh(Y)$

In this formula, μ is a step size parameter that has an effect on the speed of convergence, $\langle\cdot\rangle$ denotes the averaging operator, and $\Phi(\cdot)$ is an activation function. The hyperbolic tangent (Fig. 2)

$$\Phi(\boldsymbol{Y}) = \tanh(\eta \cdot \boldsymbol{Y}) \quad (1)$$

is widely used [1, 2] as a nonlinear activation function, where η is a scaling parameter to control the nonlinearity of Φ.

In frequency domain BSS, the signals obtained by a DFT are complex numbers. Thus, the calculations of $\Delta\boldsymbol{W}$ and $\Phi(\cdot)$ has been extended for complex numbers [5]:

$$\Delta\boldsymbol{W} = \mu\,[I - \langle\Phi(\boldsymbol{Y})\boldsymbol{Y}^H\rangle]\,\boldsymbol{W}, \quad (2)$$

$$\Phi(\boldsymbol{Y}) = \tanh[\eta \cdot re(\boldsymbol{Y})] + j \cdot \tanh[\eta \cdot im(\boldsymbol{Y})], \quad (3)$$

where $\boldsymbol{Y}^H$ represents the conjugate transpose of $\boldsymbol{Y}$, and $re(\boldsymbol{Y})$ and $im(\boldsymbol{Y})$ are the real and imaginary parts of $\boldsymbol{Y}$, respectively.

Based on (2), $\boldsymbol{W}$ converges to a point that satisfies

$$\langle\Phi(Y_p)Y_q^*\rangle = 0 \quad (p \neq q), \quad (4)$$

$$\langle\Phi(Y_p)Y_q^*\rangle = 1 \quad (p = q), \quad (5)$$

where Y_q^* is the complex conjugate of Y_q. The first equation (4) concerns the mutual independence of Y_p and Y_q. The second equation (5) makes the average amplitude of Y_p converge to some value near 1 since the range of Φ is from -1 to 1. Decomposing (5) into real and imaginary parts, we have

$$\langle\Phi[re(Y_p)]re(Y_p) + \Phi[im(Y_p)]im(Y_p)\rangle = 1 \quad (6)$$

$$\langle\Phi[im(Y_p)]re(Y_p) - \Phi[re(Y_p)]im(Y_p)\rangle = 0 \quad (7)$$

Equation (7) imposes an additional constraint that $re(Y_p)$ and $im(Y_p)$ should be mutually independent [7]. This constraint is too strong since there may be a case that $re(Y_p)$ and $im(Y_p)$ are not mutually independent.

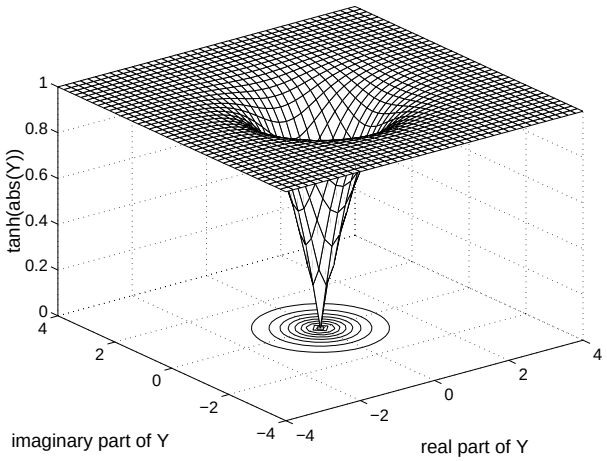

Fig. 3. Hyperbolic tangent $\tanh[abs(Y)]$ for a complex number Y

To avoid constraint (7), another formula was proposed [6, 7] and has been used in [8, 9, 13, 14]:

$$\Delta \boldsymbol{W} = \mu \left[\text{diag}(\langle \Phi(\boldsymbol{Y}) \boldsymbol{Y}^H \rangle) - \langle \Phi(\boldsymbol{Y}) \boldsymbol{Y}^H \rangle\right] \boldsymbol{W}. \quad (8)$$

According to this formula, $\boldsymbol{W}$ converges to a point that satisfies only (4), and the amplitudes of $\boldsymbol{Y}$ do not change much during ICA.

3. A NEW ACTIVATION FUNCTION

In this section, we propose a new activation function that solves the problem of constraint (7) caused by activation function (3). The new function is based on the polar coordinates of a complex number:

$$\Phi(\boldsymbol{Y}) = \tanh[\eta \cdot abs(\boldsymbol{Y})] \cdot e^{j \cdot angle(\boldsymbol{Y})}, \quad (9)$$

where $abs(\boldsymbol{Y})$ and $angle(\boldsymbol{Y})$ are the absolute values and the angles of $\boldsymbol{Y}$, respectively. It consists of two parts: amplitude part $\tanh[\eta \cdot abs(\cdot)]$ and phase part $e^{j \cdot angle(\cdot)}$. The amplitude part changes the amplitudes of $\boldsymbol{Y}$. Figure 3 shows function values of $\tanh[\eta \cdot abs(Y)]$ for a complex number Y. The phase part concerns the phases of $\Phi(\boldsymbol{Y})$, and maintains the phases equal to the phases of $\boldsymbol{Y}$. We can see that the new activation function is a natural extension of ordinary function (1), and produces the same value as (1) for a real number.

If we use this activation function, constraint (7) does not appear. Let $\theta = angle(Y_p)$. Since Y_p^* is a complex conjugate of Y_p,

$$\begin{aligned} \Phi(Y_p) Y_p^* &= \tanh[\eta \cdot abs(Y_p)] \cdot e^{j\theta} \cdot abs(Y_p) \cdot e^{-j\theta} \\ &= \tanh[\eta \cdot abs(Y_p)] \cdot abs(Y_p). \end{aligned}$$

Hence, the imaginary part of $\langle \Phi(Y_p) Y_p^* \rangle$ in (5) becomes 0.

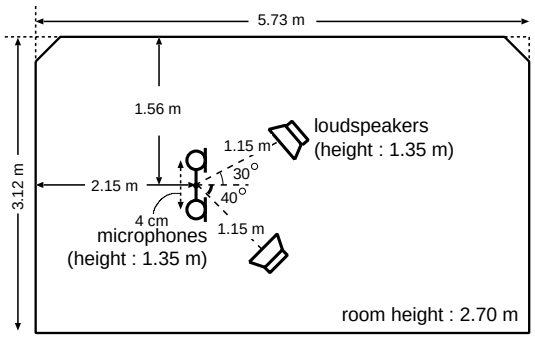

Fig. 4. Layout of a room used to record impulse responses

4. EXPERIMENTS AND DISCUSSIONS

To show the effectiveness of the new activation function, we conducted experiments to compare the performance of a BSS system using the following three combinations:

Polar-I using (9) for $\Phi(\cdot)$ and (2) for $\Delta \boldsymbol{W}$,

Cartesian-I using (3) for $\Phi(\cdot)$ and (2) for $\Delta \boldsymbol{W}$,

Cartesian-diag using (3) for $\Phi(\cdot)$ and (8) for $\Delta \boldsymbol{W}$.

4.1. Conditions for experiments and overall results

Experiments were conducted for speech signals that were convolved with impulse responses and then mixed. Figure 4 shows the layout of the room used to record the impulse responses. The numbers of sources and observations were both 2. The reverberation time of the room could be changed. We used two sets of impulse responses whose reverberation times were 150 ms and 300 ms. We selected two speech signals from the ASJ continuous speech corpus. The lengths of the speech signals were about eight seconds, and the entire eight seconds of the mixed data were used for ICA. We used different frame lengths T of a windowed DFT depending on the reverberation time as shown in Table 1. Since the sampling rate was 8000 Hz, 150 ms and 300 ms correspond to 1200 points and 2400 point, respectively.

To avoid the permutation problem [8, 9] of frequency domain BSS, we assumed that the first source signal came from the left-hand side and the other came from the right-hand side. Then, by using the technique of a null beam former, we set the initial value of $\boldsymbol{W}$ such that the first and second rows of the matrix had steep null directivity patterns towards 60° and -60°, respectively. Amplitude ambiguities were solved by the technique proposed in [8].

Table 1 shows results of SNRs for the three combinations. The numbers are the averages of SNRs at two outputs. In the column "Ref", SNRs measured with the speech sounds used in learning are shown, and SNRs measured with impulses [14] are shown in "Imp". We can see that

	$T_R = 150$ ms		$T_R = 300$ ms	
	Ref	Imp	Ref	Imp
Polar-I	18.3	19.7	12.7	16.3
Cartesian-I	17.9	19.4	12.3	15.6
Cartesian-diag	17.8	18.0	11.9	14.6

$T = 1024$ ($T_R = 150$ ms), 2048 ($T_R = 300$ ms)
$S = 256$, μ = 0.1, η = 100, #iteration = 100
Ref: measured with speech sounds
Imp: measured with impulses

Table 1. SNRs (dB) for different activation functions and different formulas for $\Delta \boldsymbol{W}$

the "Polar-I" case outperforms the others, and the results of "Cartesian-diag" are not as good as others, especially in "Imp". The reasons of these results are discussed in the next subsections.

4.2. Comparison between "Polar-I" and "Cartesian-I"

In order to see how the two $\Phi(\cdot)$ functions behave, we plot the values of $\Phi(\boldsymbol{Y})$ in Figs. 5 and 6. These data were obtained at the final learning step in the 200th (1554.7 Hz) frequency bin for the 150 ms cases. We set $\eta = 100$, which is the parameter to control the nonlinearity of $\Phi(\cdot)$.

In the case of (3), many samples were put at one of the four corners: $1 + j$, $1 - j$, $-1 + j$, $-1 - j$. On the other hand, in the case of (9), the samples were put on the unit circle $e^{j\theta}$. We can see that formula (9) represents more of the information of $\boldsymbol{Y}$, especially the phase information, than formula (3). In addition, we can say that the entropy of $\Phi(\boldsymbol{Y})$ is larger in the case of (9) than in the case of (3). This fact can be a reason why "Polar-I" outperforms "Cartesian-I".

Next, we discuss the additional constraint (7), which is imposed in the "Cartesian-I" case. Figures 7 and 8 show the absolute values and the imaginary parts of $[I - \langle \Phi(\boldsymbol{Y}) \boldsymbol{Y}^H \rangle]$, respectively, when using activation function (3). These data were obtained in the 100th (773.4 Hz) frequency bin for the 150 ms cases. In Fig. 7, we see oscillations that hinder convergence. These oscillations come from the imaginary parts of the diagonals of $[I - \langle \Phi(\boldsymbol{Y}) \boldsymbol{Y}^H \rangle]$ as shown in Fig. 8.

If we use activation function (9), we can eliminate such oscillations as discussed in Section 3. Figure 9 shows absolute values of $[I - \langle \Phi(\boldsymbol{Y}) \boldsymbol{Y}^H \rangle]$ when using (9). We can see smooth convergence. Clearly, the mutual information among $\boldsymbol{Y}$ is well minimized in this case than in the case of (3).

4.3. Comparison between "Polar-I" and "Cartesian-diag"

The main difference between the two is in the calculation of $\Delta \boldsymbol{W}$. As discussed in Section 2, formula (2) makes the average amplitude of Y_p approach some value near 1, whereas formula (8) does not change the amplitude much. The initial value of the average amplitude of Y_p differs considerably from frequency to frequency, since given signals are not generally flat in frequency.

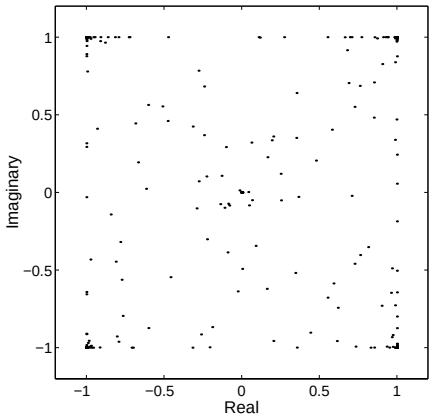

Fig. 5. Values of activation function (3)

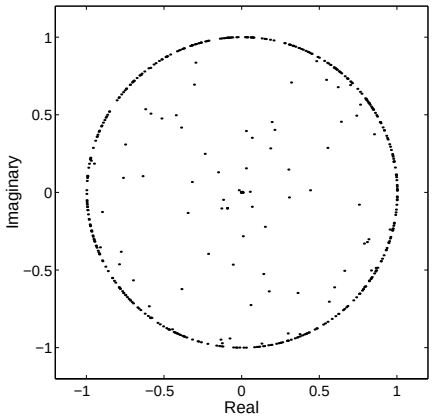

Fig. 6. Values of activation function (9)

Now, we discuss how the average amplitudes of $\boldsymbol{Y}$ affects the speed of convergence in the case of "Cartesian-diag". Since $\boldsymbol{Y}(m) = \boldsymbol{W}\boldsymbol{X}(m)$, the amount of change of $\boldsymbol{Y}$ in a learning step is

$$\begin{aligned}\Delta \boldsymbol{Y}(m) &= \Delta \boldsymbol{W}\boldsymbol{X}(m) \\ &= \mu \left[\mathrm{diag}(\langle \Phi(\boldsymbol{Y}) \boldsymbol{Y}^H \rangle) - \langle \Phi(\boldsymbol{Y}) \boldsymbol{Y}^H \rangle \right] \boldsymbol{Y}(m).\end{aligned}$$

Thus, the amount of change of each Y_p is

$$\Delta Y_p(m) = \mu \sum_{q \neq p} \langle \Phi(Y_p) Y_q^* \rangle \, Y_q(m).$$

Here, $\langle \Phi(Y_p) Y_q^* \rangle$ is approximately proportional to the average amplitude of Y_q if the mutual information between Y_p and Y_q is the same. Therefore, the average amplitude accelerates the convergence speed if it is large, and inversely, it reduces the speed if it is small.

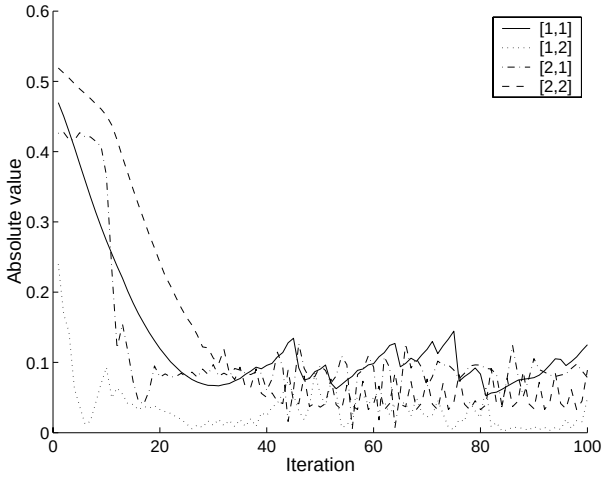

Fig. 7. Absolute values of $[I - \langle \Phi(\boldsymbol{Y})\boldsymbol{Y}^H \rangle]$ using (3)

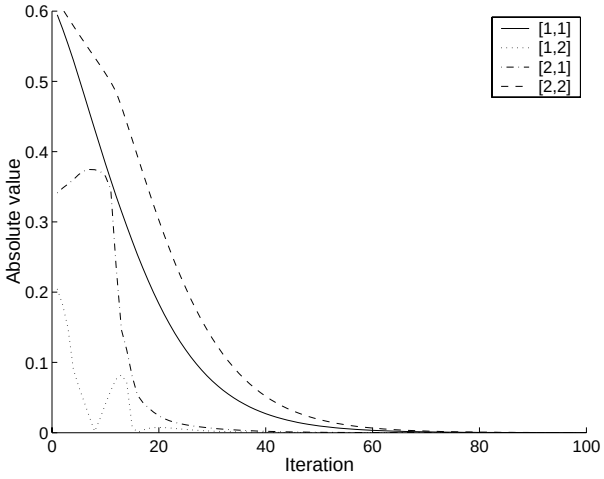

Fig. 9. Absolute values of $[I - \langle \Phi(\boldsymbol{Y})\boldsymbol{Y}^H \rangle]$ using (9)

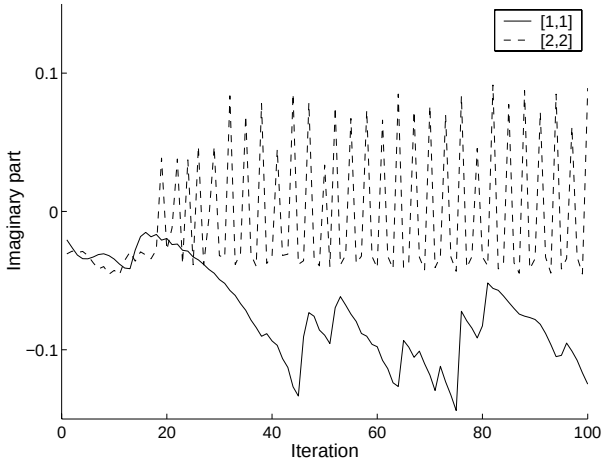

Fig. 8. Imaginary part of $[I - \langle \Phi(\boldsymbol{Y})\boldsymbol{Y}^H \rangle]$ using (3)

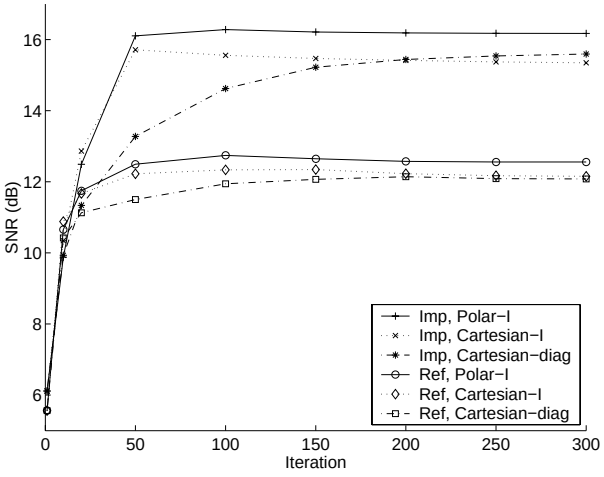

Fig. 10. Convergence speeds

Based on the discussion above, the convergence speed for each frequency differs considerably in the case of (8), whereas the speed is gradually normalized in the case of (2). Figure 10 shows convergence speeds for the 300 ms cases. We can see that more iterations are needed to reach a point of convergence in the case of "Cartesian-diag" than in the case of "Polar-I" and "Cartesian-I". To accelerate the convergence speed in "Cartesian-diag", we could set a larger value for step size μ. However, the larger μ might be too large to converge smoothly for a frequency bin with large energy. Therefore, some normalization of step size in each frequency bin would be needed for the case of "Cartesian-diag".

5. CONCLUSIONS

We have proposed a new activation function for a complex-valued ICA algorithm. The function eliminates constraint (7), and enables us to use original gradient formula (2) without oscillations that hinder convergence. Another gradient formula (8) has been used to avoid the oscillation problem. This formula, however, exhibits irregular convergence speeds among frequency bins in our experiments. We obtained improved SNRs and good convergence characteristics by using the new activation function with the original gradient formula.

REFERENCES

[1] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, no. 6, pp. 1129–1159, 1995.

[2] T. W. Lee, *Independent component analysis - Theory and applications*, Kluwer academic publishers, 1998.

[3] S. Amari, A. Cichocki, and H. H. Yang, "A new learning algorithm for blind signal separation," in *Advances in Neural Information Processing Systems*. 1996, vol. 8, pp. 757–763, The MIT Press.

[4] S. Haykin, Ed., *Unsupervised adaptive filtering*, John Wiley & Sons, 2000.

[5] P. Smaragdis, "Blind separation of convolved mixtures in the frequency domain," *Neurocomputing*, vol. 22, pp. 21–34, 1998.

[6] N. Murata and S. Ikeda, "An on-line algorithm for blind source separation on speech signals," in *Proc. Int. Symposium on Nonlinear Theory and Its Application (NOLTA '98)*, 1998, pp. 923–926.

[7] S. Kurita, H. Saruwatari, S. Kajita, K. Takeda, and F. Itakura, "Blind signal separation using directivity pattern," in *Technical Report of Japanese Society for Artificial Intelligence*, Nov. 1999, pp. 21–26.

[8] S. Ikeda and N. Murata, "A method of ICA in time–frequency domain," in *Proc. ICA '99*, Jan. 1999, pp. 365–370.

[9] S. Kurita, H. Saruwatari, S. Kajita, K. Takeda, and F. Itakura, "Evaluation of blind signal separation method using directivity pattern under reverberant conditions," in *Proc. ICASSP2000*, June 2000, pp. 3140–3143.

[10] L. Parra and C. Spence, "Convolutive blind separation of non-stationary sources," *IEEE Trans. Speech Audio Processing*, vol. 8, no. 3, pp. 320–327, May 2000.

[11] M. Z. Ikram and D. R. Morgan, "Exploring permutation inconsistency in blind separation of speech signals in a reverberant environment," in *Proc. ICASSP2000*, 2000, pp. 1041–1044.

[12] F. Asano, S. Ikeda, M. Ogawa, H. Asoh, and N. Kitawaki, "A combined approach of array processing and independent component analysis for blind separation of acoustic signals," in *Proc. ICASSP2001*, 2001, MULT-P2.1.

[13] S. Araki, S. Makino, T. Nishikawa, and H. Saruwatari, "Fundamental limitation of frequency domain blind source separation for convolutive mixture of speech," in *Proc. ICASSP2001*, 2001, MULT-P2.3.

[14] R. Mukai, S. Araki, and S. Makino, "Separation and dereverberation performance of frequency domain blind source separation for speech in a reverberant environment," in *Proc. Eurospeech2001*, Sept. 2001.