An Extension of The Herault-Jutten Network to Signals Including Delays for Blind Separation

Tatsuya Nomura* [1] , Masaki Eguchi† [2],
Hiroaki Niwamoto‡ [3], Humio Kokubo† [4], and Masayuki Miyamoto‡ [5]
* ATR Human Information Processing Research Laboratories,
2-2, Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-02, Japan
† Energy Conversion Laboratories, SHARP Corp. , 282-1, Hajikami,
Shinjo-cho, Kitakatsuragi-gun, Nara 639-21, Japan
‡ Information Technology Research Laboratories, SHARP Corp. ,
2613-1, Ichinomoto-cho, Tenri, Nara 632, Japan
E-mail: [1] nomura@hip.atr.co.jp, [2] eguchi@shinjo.sharp.co.jp,
[3] niwamoto@shpcsl.sharp.co.jp, [4] kokubo@shinjo.sharp.co.jp,
and [5] miyamoto@shpcsl.sharp.co.jp

Abstract: Many neural networks for the blind separation problem have recently been proposed. However, in most of them, delays on input signals to networks are not considered. In this paper, we propose an extension of the Herault-Jutten network that is applied to input signals including delays. Moreover, we present results of comparative simulations between the original Herault-Jutten network and our method for cases where the input signals are direct signals from sources and delay signals, such as reflections off walls.

1. Introduction

The problem of multi-channel blind separation of sources such as the "cocktail-party" problem arises in diverse fields in neural computation (including the hearing and olfactory systems) and in applied science (including radar, speech processing, and digital communications). This problem is how to separate source signals from observable signals in which the sources are mixed through an unknown channel. The problem was formalized by Jutten and Herault in the 1980's [1] and many neural network models for this problem have recently been proposed [3][4][5][6][7].

In much research for the blind separation problem, the following formalization is basically used. $X_1(t), X_2(t), \ldots, X_n(t)$ $(t = 0, 1, \ldots)$ are source signals and are assumed stationary and statistically independent of one another. When $E_1(t), E_2(t), \ldots, E_n(t)$ are observable signals, the following relation is assumed:

$$E(t) = AX(t) \quad (t = 0, 1, \ldots) \tag{1}$$
$$X(t) = (X_1(t), X_2(t), \ldots, X_n(t))^t, \; E(t) = (E_1(t), E_2(t), \ldots, E_n(t))^t$$

where, A is an $n \times n$ non-singular mixing matrix with constant values and $(\cdot)^t$ denotes the transpose of a vector. The purpose of the blind separation problem is to estimate the unknown matrix A using only the observable signals $E_1(t), E_2(t), \ldots, E_n(t)$ and to find the inverse operation that separates the source signals from the observable signals.

However, the above formalization does not reflect some of the problems in the real world. For example, let us imagine the situation where two persons speak in a closed room; now, consider separating of each speech signal from the outputs of two microphones. Although it is assumed that each direct speech signal reaches the two microphones at the same time in Equation (1), in fact, each signal reaches the microphones at a different time. Furthermore, reflection signals from the walls, floor, and ceiling reach the microphones with greater delay. Therefore, the following relation between source signals and observational signals exists:

$$E(t) = \sum_{i=0}^{m} A(i)X(t-i) \quad (t = 0, 1, \ldots) \tag{2}$$

where, the definitions and the assumptions of $X(t)$ and $E(t)$ are the same as those in Equation (1) and $A(i)$ $(i = 0, \ldots, m)$ are $n \times n$ matrices with constant values. Methods using the formula in Equation (1), such as the Herault-Jutten network (H-J network)[1], are unable to separate source signals from observable signals like in Equation (2).

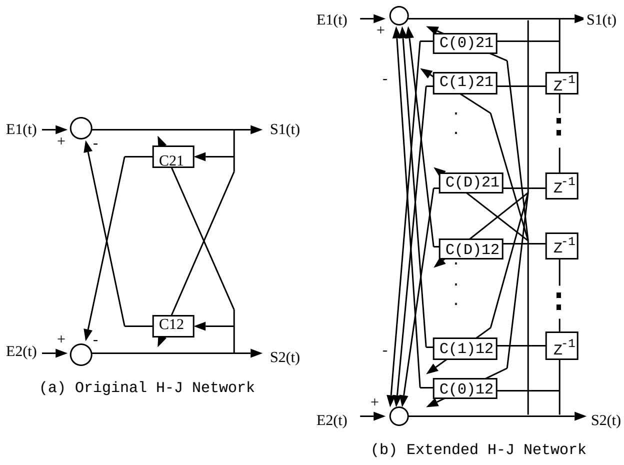
Matsuoka and Kawamoto have proposed a neural network appropriate for the observational signals in Equation (2) [7]. In this paper, we extend the H-J network and propose a neural network model for blind separation where observable signals are made from delayed source signals. Moreover, we present results of comparative simulations between our method and the original H-J network.

2. Herault-Jutten Network and its Extension

2.1 Feedforward Process

Figure 1 shows the original H-J network and the extension we propose for the case of $n = 2$.

When the observable signal vector $E(t)$ is given as an input, the n-dimensional output vector $S(t) = (S_1(t), S_2(t), \ldots, S_n(t))$ of the H-J network



(a) Original H-J Network

(b) Extended H-J Network

Figure 1: Original and Extended H-J Networks

is given by

$$S(t) = E(t) - CS(t) = (I + C)^{-1}E(t) \tag{3}$$

where, $C = (C_{ij})$ are $n \times n$ matrices with $C_{ii} = 0$ $(i = 1, \ldots, n)$. We consider the effect of delayed signals and extend the H-J network to the following formula:

$$\begin{aligned} S(t) &= E(t) - \sum_{k=0}^{D} C(k)S(t-k) \\ &= (I + C(0))^{-1}\left(E(t) - \sum_{k=1}^{D} C(k)S(t-k)\right) \end{aligned} \tag{4}$$

where, $C(k) = (C(k)_{ij})$ $(k = 0, \ldots, D)$ are $n \times n$ matrices with $C(k)_{ii} = 0$ $(i = 1, \ldots, n, k = 0, \ldots, D)$. From the definition, if D equals 0, the feedforwrd process of this network equals that of the original H-J network.

2.2 Learning Rule

In the original H-J network, the weights are updated based on the gradient descent method for the function $S_i(t)^2$ of $C_{ij} (j = 1, \ldots, n, i \neq j)$. From Equation (3),

$$\frac{\partial S(t)}{\partial C_{ij}} = -(I + C)^{-1}\frac{\partial C}{\partial C_{ij}}S(t) \tag{5}$$

is given. From the condition that the matrix $\frac{\partial C}{\partial C_{ij}}$ has 1 at the (i, j) part and 0 at the other part and the first order expansion of $(I + C)^{-1}$, the following learning rule is derived:

$$\frac{dC_{ij}}{dt} = \alpha S_i(t) S_j(t) \quad (i, j = 1, \ldots, n, i \neq j) \tag{6}$$

where, α is the learning parameter. As a result, when the weights converge, the output signals are independent of one another.

We also derive a learning rule from the gradient descent method for the function $S_i(t)^2$ of $C(k)_{ij}(j = 1, \ldots, n, i \neq j, k = 0, \ldots, D)$. With Equation (4),

$$\frac{\partial S(t)}{\partial C(0)_{ij}} = -(I + C(0))^{-1} \left\{ \frac{\partial (I + C(0))}{\partial C(0)_{ij}} S(t) + \sum_{l=1}^{D} C(l) \frac{\partial S(t-l)}{\partial C(0)_{ij}} \right\} \tag{7}$$

is given. Moreover, for $k > 0$,

$$\frac{\partial S(t)}{\partial C(k)_{ij}} = -(I + C(0))^{-1} \left\{ \frac{\partial C(k)}{\partial C(k)_{ij}} S(t-k) + \sum_{l=1}^{D} C(l) \frac{\partial S(t-l)}{\partial C(k)_{ij}} \right\} \tag{8}$$

is given. Here, we regard $S(t-l)$ $(l > 0)$ as a constant for $C(k)_{ij}$. From the condition that matrices $\frac{\partial(I+C(0))}{\partial C(0)_{ij}}$ and $\frac{\partial C(k)}{\partial C(K)_{ij}}$ have 1 at the (i, j) part and 0 at the other part and the first order expansion of $(I + C(0))^{-1}$, we obtain the following learning rule:

$$\frac{dC(k)_{ij}}{dt} = \alpha S_i(t) S_j(t-k) \quad (i, j = 1, \ldots, n, i \neq j, k = 0, \ldots, D) \tag{9}$$

When the weights converge, the output signals are independent of one another in the same way as in the original H-J network. This learning rule is an extension of that of the original H-J network.

Note:

Although we include the matrix $C(0)$ in the definition of our network, we often set $C(0) = 0$. Therefore, in the strict sense, our network is not an extension of the H-J network. However, using the simulations, it was shown that the existence of $C(0)$ did not affect the separation ability of our network.

3. Simulations

We executed comparative simulations between the original H-J network and our neural network for the observable signals in Equation (2). As shown in Figure 2, we assumed that auditory signals from two sources were mixed and reached two microphones far from the sources ($n = 2$).

Here, we assumed that the first (resp. the second) microphone was placed in front of the first (resp. the second) source and that the line between the two sources and the line between the two microphones were perpendicular to the wall.

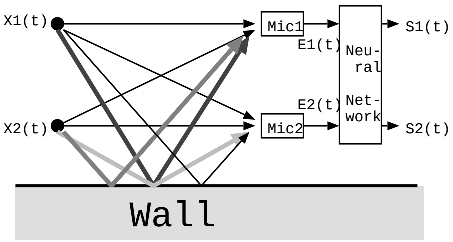


Figure 2: The situation assumed in the simulations

3.1 Observable Signals for Experiments

We used the following two kinds of signals as source signals:

$$\left.\begin{array}{rcl} X_1(t) & = & \sin(2\pi t/100) \\ X_2(t) & = & \text{a random noise with amplitude 2.0} \end{array}\right\} \qquad (10)$$

Figure 3 shows these source signals.

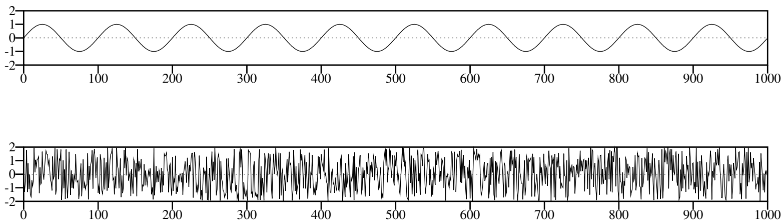


Figure 3: Source Signals for the Simulations (Upper: $X_1(t)$, Lower: $X_2(t)$)

Moreover, we set the mixture matrices $A(i)$ in Equation (2) for the following two cases:

Case 1: Reflection signals from the wall did not exist:

$$D = 3,\ A(0) = A(2) = 0,\quad A(1) = \begin{pmatrix} 0.7 & 0.0 \\ 0.0 & 0.7 \end{pmatrix},\quad A(3) = \begin{pmatrix} 0.0 & 0.3 \\ 0.3 & 0.0 \end{pmatrix}$$

Case 2: Reflection signals from the wall existed:

$$D = 7,\quad A(0) = A(2) = A(4) = 0,\quad A(1) = \begin{pmatrix} 0.7 & 0.0 \\ 0.0 & 0.7 \end{pmatrix},$$

$$A(3) = \begin{pmatrix} 0.0 & 0.3 \\ 0.3 & 0.0 \end{pmatrix},\quad A(5) = \begin{pmatrix} 0.0 & 0.18 \\ 0.18 & 0.14 \end{pmatrix},\quad A(7) = \begin{pmatrix} 0.04 & 0.0 \\ 0.0 & 0.0 \end{pmatrix}$$

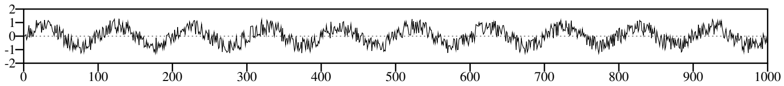


Figure 4: Observable Signals for the Simulations in Case 1 (not including delayed reflection signals, Upper: $E_1(t)$, Lower: $E_2(t)$)

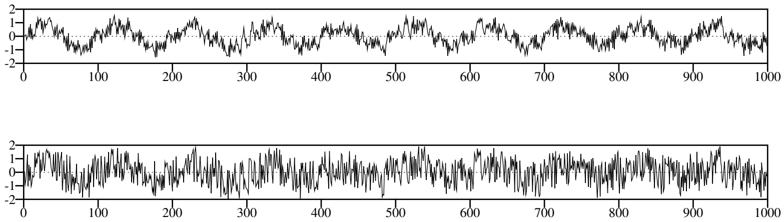


Figure 5: Observable Signals for the Simulations in Case 2 (not including delayed reflection signals, Upper: $E_1(t)$, Lower: $E_2(t)$)

Figures 4 and 5 show these observable signals. These matrices were calculated based on the following assumptions; the distance between each source and the front microphone and the distance between a source and the wall corresponded to one sampling time, the decay rate of the signal from each source to the front (resp. non-front) microphone was 0.7 (resp. 0.3) and the decay rates of signals were inversely proportional to the distance rates.

3.2 Results of Simulations

In the simulations, we set the learning parameter $\alpha = 0.01$ in both the original and the extended H-J network. Moreover, we prepared 5 delay units for Case 1 and 10 delay units for Case 2 in the extended H-J network.

Figures 6, 7, and 8 show the outputs of the original H-J network, our extended H-J network, and that with $C(0) = 0$ for the observable signals in Case 1. Moreover, 9, 10, and 11 show the outputs of them for the observable signals in Case 2. We show only the signal $S_1(t)$ for each method because the signal $S_2(t)$ became almost random in these simulations.

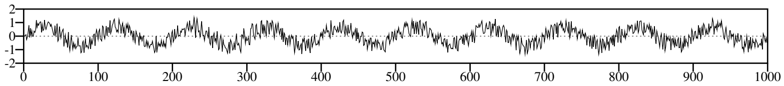


Figure 6: Output Signal $S_1(t)$ of the Original H-J Network for the Observable Signals in Case 1

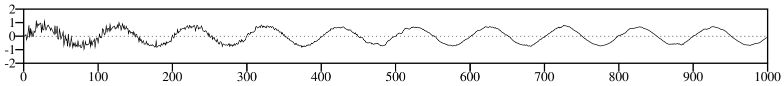


Figure 7: Output Signal $S_1(t)$ of the Extended H-J Network for the Observable Signals in Case 1 ($D = 5$, $C(0)$ existed)

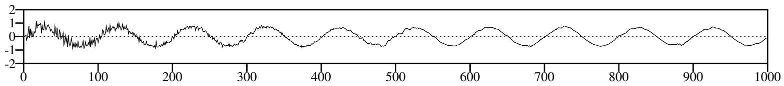


Figure 8: Output Signal $S_1(t)$ of the Extended H-J Network for the Observable Signals in Case 1 ($D = 5$, $C(0)$ did not exist)

As shown in Figures 6 and 9, the original H-J network was not able to separate the source signals because of the effect of the delayed signals included in the observable signals. On the other hand, for our extended H-J networks, the output signals similar to the original signals were obtained after about $t = 500$. Although the envelopes on the waves of the output signals were a little distorted, the output signals of our networks were much more similar to the source signal (the sine wave) than that of the original H-J network, regardless of the existence of $C(0)$.

Moreover, we regarded these networks as echo cancellers and evaluated "Echo Return Loss Enhancement (ERLE)", often used in the evaluation for echo cancellers. In these simulations, we defined ERLE with time average in the following:

$$ERLE(t) = 10 \log_{10} \left(\frac{E[(0.7X_1(t-1) - S_1(t))^2]}{E[(0.7X_1(t-1))^2]} \right) \ [db]$$

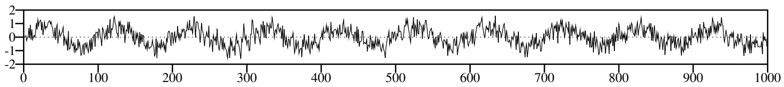


Figure 9: Output Signal $S_1(t)$ of the Original H-J Network for the Observable Signals in Case 2

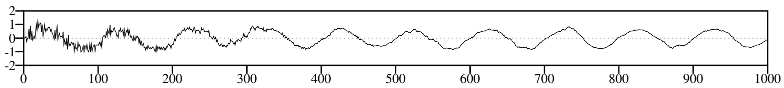


Figure 10: Output Signal $S_1(t)$ of the Extended H-J Network for the Observable Signals in Case 2 ($D = 10$, $C(0)$ existed)

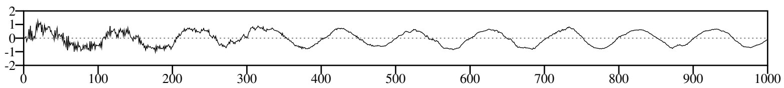


Figure 11: Output Signal $S_1(t)$ of the Extended H-J Network for the Observable Signals in Case 2 ($D = 10$, $C(0)$ did not exist)

Here, $E[\cdot]$ is a expected value of a stochastic variable. We compared $S_1(t)$ with $0.7X_1(t-1)$ because the observable signals in Case 1 include the source signal $X_1(t)$ with the maximum amplitude 0.7 and shortest deley 1. Moreover, in the acutual calculations of the above expected values, we used time average values instead of the real expected values. Figure 12 shows ERLE of each method. The ERLE of the original H-J network remained about at $-5db$ all the time. In contrast, the ERLEs of our networks decreased below $-20db$ after about $t = 500$ and our networks showed the higher capacity of signal separation than that of the original H-J network in cases where observable signals includes delayed source signals.

Table 1 shows the weights finally obtained in our network. The weights corresponding to $C(0)$ were almost 0 and did not affect the outputs. Moreover, the weights corresponding to the second delay unit were bigger than the others. A possible reason is that the delay of the earliest source signal included in the observable signals was 1, that of the latest one was 3, and the difference

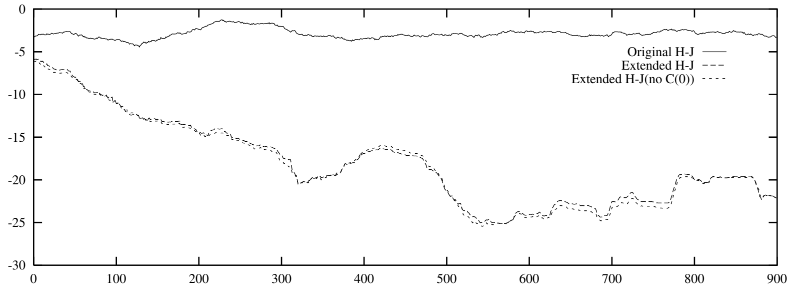


Figure 12: ERLE on $S_1(t)$ on each method for the observable

between them was 2 for Case 1.

Table 1: The Weights Finally Obtained in the Extended J-H network for Case 1

	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
$C(k)_{12}$	0.00	-0.01	0.42	0.00	-0.01	-0.01
$C(k)_{21}$	0.00	0.09	0.08	0.06	0.04	0.07

The above results are ones for simple examples. We have executed simulations for mixed real and auditory signals, based on an assumption similar to that of the above simulations. This will be dealt with in future works.

3.3 Discussion

Using the above simulations, we verified the effectiveness of our network to some degree. However, there remains several problems.

Although we could prepare a sufficient number of delay units in our network for the above simulations, the number of delay units is limited in a real environment. As a solution, we should consider investigating observable signals, inferring the numbers of delayed source signals included in observable signals, and preparing only the units corresponding to them. Moreover, we should consider adaptively adjusting the corresponding delay numbers within finite delay units given in advance.

We also have another important problem. For the original H-J network, the existence of optimal solutions for signal separation is guaranteed and the conditions under which the learning of the network reaches the optimal solutions have been analyzed [1][2]. We should analyze the existence of optimal solutions for the separation of signals in Equation (2) and the condition of successful learning in Equation (9).

4. Conclusion

We have proposed an extension of the H-J network to cope with signals including delays and have verified the effectiveness of our network by comparative simulations with the original H-J network. It was concluded that our extended H-J network was superior to the original H-J network for signals including delays.

References

[1] C. Jutten and J. Herault, "Blind separation of sources, Part I : An adaptive algorithm based on neuromimetic architecture," **Signal Processing**, vol. 24, pp. 1–10, 1991.

[2] E. Sorouchyari, "Blind separation of sources, Part III : Stability analysis," **Signal Processing**, vol. 24, pp. 21–29, 1991.

[3] L. Molgedey and H. G. Schuster, "Separation of a Mixture of Independent Signals Using Time Delayed Correlations," **PHISICAL REVIEW LETTERS**, vol. 72, no. 23, pp. 3634–3637, 1994.

[4] S. Amari, A. Cichoski, and H. H. Yang, "Recurrent Neural Networks For Blind Separation Of Sources," in **Proc. 1995 International Workshop on Nonlinear Theory and Its Applications (NOLTA'95)**, pp. 37–42, Dec 1995.

[5] A. J. Bell and T. J. Sejnowski, "Fast blind separation based on information theory," in **Proc. 1995 International Workshop on Nonlinear Theory and Its Applications (NOLTA'95)**, pp. 43–47, Dec 1995.

[6] A. Cichoski, W. Kasprzak, and S. Amari, "Multi-Layer Neural Networks with a Local Adaptive Learning Rule for Blind Separation of Sources Signals," in **Proc. 1995 International Workshop on Nonlinear Theory and Its Applications (NOLTA'95)**, pp. 61–65, Dec 1995.

[7] K. Matsuoka and M. Kawamoto, "Blind Signal Separation Based on a Mutual Information Criterion," in **Proc. 1995 International Workshop on Nonlinear Theory and Its Applications (NOLTA'95)**, pp. 85–90, Dec 1995.