Estimation of Output Disturbance in Auto-Regressive Model via Independent Component Analysis

R. Tanaka, K. Kawaguchi, J. Endo, H. Shibasaki, Y. Hikichi, and Y. Ishida

***Abstract*—This paper explains and demonstrates how to estimate an output disturbance in an auto-regressive model. This method uses the independent component analysis (ICA) technique, which restores source signals from their linear mixtures under the assumption that the source signals are mutually independent. The estimation is achieved by a model whose source signals consist of input and output disturbance, and observed signals consist of input and output. To solve the ICA problem, a natural gradient method based on mutual information is adopted. As a result, in this simulation, the NRR of our proposed method shows an improvement of about 4.0 [dB] compared with that of a conventional method.**

***Index Terms*—Independent Component Analysis (ICA), Blind Signal Separation (BSS), kullback-leibler divergence, Auto Regressive (AR) model.**

I. Introduction

The goal of blind signal separation (BSS) is to restore source signals from their mixtures without *priori*-knowledge of the source signals and a mixing process which generates the mixtures. In the case of the instantaneous mixture model, in which the mixtures are generated as linear combinations of the source signals, ICA [1] is one of the successful approaches. ICA has been applied in various areas such as speech, audio, images, and communications. In ICA, the observed signals, which are outputs from a mixing process, are linearly transformed such that separated signals, which are outputs from separating process, are statistically as independent as possible under the assumption that the source signals are mutually independent.

Various control system techniques to reduce the disturbance have been reported [2], [3]; however, most of the methods are required an information of accurate plant model parameter. From such a background, if a control system is a stable plant and the input signals are mutually independent, ICA technique makes it possible to estimate an output disturbance without depending on a system transfer function.

This paper describes an estimation process based on an ICA algorithm in an single-input single-output (SISO) control system. The estimation is achieved by a model whose source signals consist of input and output disturbance, and observed signals consist of input and output.

II. BSS Problem for ICA

It is assumed that source signals $s_n(t)$ consist of N signal sources described by $s_n(t)$ $(n=1,\dots,N)$, and observed signals consist of an M-channel sensor array $(M \ge N)$ described by $x_m(t)$ $(m=1,\dots,M)$, where t indicates a discrete-time index. The BSS model is formulated as

$$x_m(t) = \sum_{n=1}^{N} a_{nm} s_n(t), \tag{1}$$

where a_{nm} is an attenuation coefficient representing a path from the nth source to the mth sensor. By using a vector or matrix notation, (1) is rewritten as

$$\boldsymbol{x}(t) = \boldsymbol{A}\boldsymbol{s}(t), \tag{2}$$

where $\boldsymbol{A} \in \mathbb{R}^{M\times N}$ is a mixing matrix, $\boldsymbol{s}(t) = [s_1(t),\dots,s_N(t)]^T$ is a source signal vector, $\boldsymbol{x}(t) = [x_1(t),\dots,x_M(t)]^T$ is an observed signal vector, and the superscript $\bullet^T$ denotes a transpose of a matrix.

On the other hand, the separating process is formulated with separated signals denoted by $\tilde{s}_n(t)$ $(n=1,\dots,N)$. The separating process is expressed by

$$\tilde{s}_n(t) = \sum_{m=1}^{M} w_{nm} x_m(t), \tag{3}$$

where w_{nm} is a coefficient for separation. By using a vector or matrix notation, (3) is also rewritten as

$$\tilde{\boldsymbol{s}}(t) = \boldsymbol{W}\boldsymbol{x}(t), \tag{4}$$

where $\tilde{\boldsymbol{s}}(t) = [\tilde{s}_1(t),\dots,\tilde{s}_N(t)]^T$ is a separated signal vector, and $\boldsymbol{W} \in \mathbb{R}^{N\times M}$ is a separating matrix.

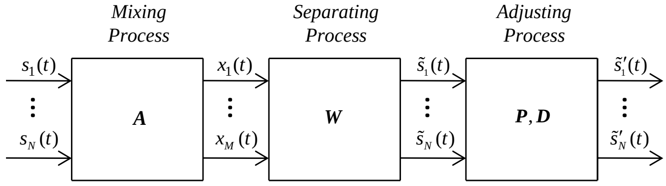

Fig. 1. Block diagram of BSS based on ICA $(M \ge N)$.

Fig. 1 shows a block diagram of BSS based on ICA. If the number of observed signals M is equal to or more than that of source signals N, and all independent components are non-Gaussian with the exception that a one Gaussian component is permitted [1], ICA estimates source signals that exclude permutation and amplitude ambiguity without an information of mixing process [4]. Hence, $\boldsymbol{W}$ satisfied by

Manuscript received September 31, 2012; revised December 31, 2012.
The authors are with School of Science and Technology, Meiji University, Kawasaki-shi, Kanagawa, 214-8571 Japan. (e-mail: rtanaka@meiji.ac.jp).

DOI: 10.7763/IJMO.2013.V3.236

$$\boldsymbol{WA} = \boldsymbol{PD} \tag{5}$$

has to be calculated, where $\boldsymbol{P}$ is a permutation matrix given by permuting row (or column) vectors of the identity matrix, and $\boldsymbol{D}$ is an amplitude matrix given by arbitrary diagonal matrix.

III. Model Formulation

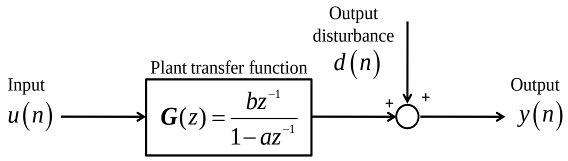

Fig. 2. Block diagram in an SISO system.

Fig. 2 shows a block diagram in SISO system with an auto-regressive model. In Fig. 2, $u(n)$ is an input, $d(n)$ is an output disturbance, $y(n)$ is an output, and $\boldsymbol{G}(z)$ is a plant transfer function as follows.

$$\boldsymbol{G}(z) = \frac{bz^{-1}}{1-az^{-1}}, \tag{6}$$

where the parameters $|a| < 1$ and b are arbitrary real values. By using (6), the relation between $u(n)$ and $y(n)$ in Z-domain is expressed by

$$Y(z) = G(z)U(z) + D(z), \tag{7}$$

where $U(z)$, $D(z)$, and $Y(z)$ denote Z-transform of $u(n)$, $d(n)$, and $y(n)$, respectively. Eq. (7) is transformed by an inverse Z-transform such that

$$\begin{aligned} y(n) &= b\sum_{k=1}^{n-1} a^{n-k-1}u(k) + d(n) \\ &\simeq b\sum_{k=1}^{\tau} a^{k}u(k) + d(n). \end{aligned} \tag{8}$$

By using (8), a matrix representation that regards $u(n)$ and $y(n)$ as observed signals is expressed by

$$\begin{bmatrix} \boldsymbol{u}(n) \\ y(n) \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \boldsymbol{\theta} & 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{u}(n) \\ d(n) \end{bmatrix}, \tag{9}$$

where $\boldsymbol{u}(n) = [u(n-1), \dots, u(n-\tau)]^T \in \mathbb{R}^{\tau\times 1}$ is an input signal vector that consists of lag elements of $u(n)$, $\mathbf{I} \in \mathbb{R}^{\tau\times\tau}$ is a unit matrix. $\boldsymbol{\theta} \in \mathbb{R}^{1\times\tau}$ is a model parameter vector, and $\tau(\le k)$ is tap length. Eq. (9) is rewritten as

$$\boldsymbol{x}' = \boldsymbol{A}'\boldsymbol{s}', \tag{10}$$

where $\boldsymbol{x}' = [\boldsymbol{u}(n), y(n)]^T \in \mathbb{R}^{(\tau+1)\times 1}$ is an observed signal vector, $\boldsymbol{s}' = [\boldsymbol{u}(n), d(n)]^T \in \mathbb{R}^{(\tau+1)\times 1}$ is a source signal vector, and $\boldsymbol{A}' \in \mathbb{R}^{(\tau+1)\times(\tau+1)}$ is a mixing matrix such that

$$\boldsymbol{A}' = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \boldsymbol{\theta} & 1 \end{bmatrix}. \tag{11}$$

By transferring (11), the problem is reduced to an ICA problem. That is, the problem is to estimate an optimal linear parameter vector $\boldsymbol{\theta}$.

IV. Conventional ICA Algorithm

Minimizing the mutual information of separated signals [5] has been adopted for an ICA algorithm. In probability distribution, this algorithm is based on a property that a joint probability density function (p.d.f.) for source signals is equal to the multiplication of each marginal p.d.f. if these signals are mutually independent. Kullback–Leibler divergence (KLD) is adopted as an independent index. KLD $J(\tilde{\boldsymbol{s}};\boldsymbol{W})$ for separating matrix $\boldsymbol{W}$ and separated signal $\tilde{\boldsymbol{s}}$ is defined as

$$\begin{aligned} J(\tilde{\boldsymbol{s}};\boldsymbol{W}) &= \int p_{\tilde{s}}(\tilde{\boldsymbol{s}};\boldsymbol{W}) \log \frac{p_{\tilde{s}}(\tilde{\boldsymbol{s}};\boldsymbol{W})}{\prod_{n=1}^{N} p_{\tilde{s}_n}(\tilde{s}_n;\boldsymbol{W})} d\tilde{s}_n \\ &= -H(\tilde{\boldsymbol{s}};\boldsymbol{W}) + \sum_{n=1}^{N} H(\tilde{s}_n;\boldsymbol{W}), \end{aligned} \tag{12}$$

where $p_{\tilde{s}}(\tilde{\boldsymbol{s}};\boldsymbol{W})$ is the joint p.d.f. of $\tilde{\boldsymbol{s}}$ determined by $\boldsymbol{W}$, and $H(\tilde{\boldsymbol{s}};\boldsymbol{W})$ is an information entropy defined as

$$\begin{aligned} H(\tilde{\boldsymbol{s}};\boldsymbol{W}) &= E[-\log p_{\tilde{s}}(\tilde{\boldsymbol{s}};\boldsymbol{W})] \\ &= -\int p_{\tilde{s}}(\tilde{\boldsymbol{s}};\boldsymbol{W}) \log p_{\tilde{s}}(\tilde{\boldsymbol{s}};\boldsymbol{W}) d\tilde{\boldsymbol{s}} \\ &= H(\boldsymbol{x}) + \log|\det \boldsymbol{W}|, \end{aligned} \tag{13}$$

where $\det\boldsymbol{W}$ is a determinant of $\boldsymbol{W}$. Hence, Eq. (12) is rewritten as

$$J(\tilde{\boldsymbol{s}};\boldsymbol{W}) = \sum_{n=1}^{N} H(\tilde{s}_n;\boldsymbol{W}) - H(\boldsymbol{x}) - \log|\det\boldsymbol{W}|. \tag{14}$$

If the components of $\tilde{\boldsymbol{s}}$ are mutually independent, the equal sign in (14) is approved. Hence, it only has to decide $\boldsymbol{W}$ that decreases $J(\tilde{\boldsymbol{s}};\boldsymbol{W})$. The partial differential of $J(\tilde{\boldsymbol{s}};\boldsymbol{W})$ for $\boldsymbol{W}$ is calculated by

$$\frac{\partial}{\partial \boldsymbol{W}} J(\tilde{\boldsymbol{s}};\boldsymbol{W}) = (E_{p_{\tilde{s}}}[\phi(\tilde{\boldsymbol{s}})\tilde{\boldsymbol{s}}^T] - \boldsymbol{I})(\boldsymbol{W})^{-T}, \tag{15}$$

where $\phi(\tilde{\boldsymbol{s}}) = [\phi(\tilde{s}_1), \dots, \phi(\tilde{s}_N)]^T$ is a score function. $\phi(\tilde{s}_n)$ is given by a differential of $\log p(\tilde{s}_n)$ like

$$\phi(\tilde{s}_n) = -\frac{d}{d\tilde{s}_n} \log p(\tilde{s}_n). \tag{16}$$

This substitutes variety of nonlinear functions $\phi(\tilde{s}_n)$ for $p(\boldsymbol{s})$ because $\boldsymbol{s}$ is unknown. To decide on its substitution, a normalized kurtosis κ is introduced such that

$$\kappa = \frac{E[\tilde{s}_n{}^4]}{E[\tilde{s}_n{}^2]} - 3, \tag{17}$$

which quantifies a sharpness of p.d.f. of signals. Based on (17), $\phi(\tilde{s}_n)$ determines whether to use $\tanh(\tilde{s}_n)$ or $\tilde{s}_n^{\,3}$ by the positive and negative values of κ such that

$$\phi(\tilde{s}_n)=\begin{cases}\tanh(\tilde{s}_n)\ , \kappa>0\\ \tilde{s}_n^{\,3}\quad\ , \kappa<0\end{cases}. \tag{18}$$

Amari et al. pointed out that (15) does not always give a steepest descent direction because the space that probability distributions make becomes a Riemann space [6]. Thus, they proposed an update rule with natural gradient such that

$$\frac{\partial}{\partial \boldsymbol{W}}J(\tilde{\boldsymbol{s}};\boldsymbol{W})=(\boldsymbol{I}-E_{p_{\tilde{s}}}[\phi(\tilde{\boldsymbol{s}})\tilde{\boldsymbol{s}}^T])\boldsymbol{W}. \tag{19}$$

V. Rearrange Update Rules

The model described in section 3 cannot be directly applied in conventional ICA techniques because $\boldsymbol{A}'$ has a special structure. Another factor is that both the terms in (9) include the same component $\boldsymbol{u}(n)$, and every element is not mutually independent. Hence, the separating matrix also requires a special structure. The entire separating matrix $\boldsymbol{W}'$ need not be updated because $\boldsymbol{A}'$ has a special structure in which each component of $\boldsymbol{A}'$ other than $\boldsymbol{\theta}$ is constant. Hence, Eq. (19) is rewritten as

$$\Delta\boldsymbol{W}=\boldsymbol{\Psi}(\boldsymbol{I}-E_{p_{\tilde{s}}}[\phi(\tilde{\boldsymbol{s}})\tilde{\boldsymbol{s}}^T])\boldsymbol{W}', \tag{20}$$

where $\boldsymbol{\Psi}\in\mathbb{R}^{(\tau+1)\times(\tau+1)}$ is a projection matrix consisting of

$$\boldsymbol{\Psi}=\mathrm{diag}(0,\ldots,0,1). \tag{21}$$

When calculating $E[\bullet]$, the sample average is substituted with the time average to lower the estimation cost. Eq. (20) makes it possible to update only $\boldsymbol{\theta}$. After (20), $\boldsymbol{W}'$ is updated as

$$\boldsymbol{W}'\leftarrow\boldsymbol{W}'+\frac{\eta\Delta\boldsymbol{W}'}{\|\Delta\boldsymbol{W}'\|_F}, \tag{22}$$

where η is an arbitrary positive step size parameter, and $\|\bullet\|_F$ denotes a Frobenius norm. By using (22), separated signal vector $\tilde{\boldsymbol{s}}'=[\boldsymbol{u}(n),d'(n)]^T\in\mathbb{R}^{(\tau+1)\times 1}$ is obtained as

$$\tilde{\boldsymbol{s}}'=\boldsymbol{W}'\boldsymbol{x}', \tag{23}$$

where $d'(n)$ is an estimated output disturbance by ICA. After ICA, $\boldsymbol{W}'$ is normalized so that all its diagonal elements are 1, because all the diagonal elements of $\boldsymbol{A}'^{-1}$ are 1.

VI. Simulations

In the preceding section, the algorithm for an output disturbance estimation in an SISO system has been shown. In this section, two simulations to evaluate the performance of signal estimation are demonstrated. The first is our proposed method. The second is Miyaura's method [7], whose algorithm adopts FastICA [8] and neural networks for ICA [9]; however, the second method is not discussed in detail in this paper owing to space constraints. The sampling time and the signal length for this control system are 0.01 sec and 5.00 sec, respectively. The system transfer function $\boldsymbol{G}(z)$ is set to two patterns like

$$\boldsymbol{G}_1(z)=\frac{z^{-1}}{1+0.3z^{-1}} \tag{24}$$

and

$$\boldsymbol{G}_2(z)=\frac{z^{-1}}{1+0.2z^{-1}+0.15z^{-2}}. \tag{25}$$

An initial parameter matrix $\boldsymbol{W}_0$ for separating matrix $\boldsymbol{W}'$ is determined by

$$\boldsymbol{W}_0=\mathbf{I}\in\mathbb{R}^{(\tau+1)\times(\tau+1)}. \tag{26}$$

$\boldsymbol{W}_0$ derives the fact that all the diagonal elements of $\boldsymbol{A}'^{-1}$ are 1 and the tap length τ is 6 points. The objective measurement in terms of noise reduction rate (NRR) is considered. NRR among $d(n)$, $y(n)$, and an estimated output disturbance $d'(n)$ is defined as

$$\mathrm{NRR[dB]}=10\log_{10}\frac{\mathrm{var}[y(n)-d(n)]}{\mathrm{var}[d'(n)-d(n)]}, \tag{27}$$

where var[•] denotes the variance of a random variable. NRR is an index that shows how many noises can be reduced from the observation point. Figure 3 shows an input $u(n)$, an output disturbance $d(n)$, and an output $y(n)$. The input $u(n)$ is a step signal, and its rise time is set to 1.00 sec. $d(n)$ is white Gaussian noise.

Fig. 4 and Fig. 5 show an estimation error in the first-order system and the second-order system, respectively. And table 1 and 2 show the result of NRR and SNR with Miyaura's method and proposed method, respectively. It is shown that our proposed method allows for more accurate estimates than Miyaura's method in objective measurement.

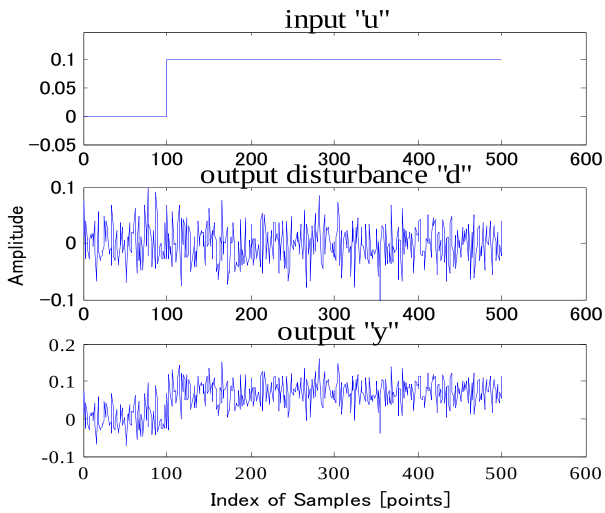

Fig. 3. Input, output disturbance, and output in simulation.

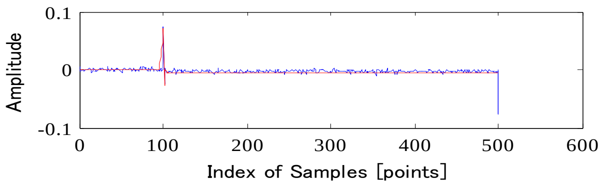

Fig. 4. Estimation error in primary system
(Blue is an conventional method, and Red line is our proposed method.)

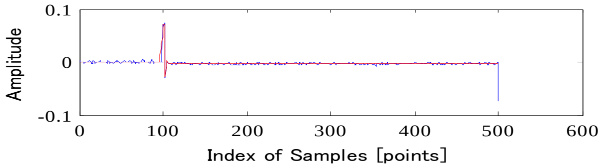

Fig. 5. Estimation error in secondary system
(Blue is an conventional method, and Red line is our proposed method.)

TABLE I: NRR AND SNR WITH CONVENTIONAL METHOD

	NRR [dB]	SNR_{out} [dB]
Conventional method ($\boldsymbol{G}_1(z)$)	13.4492	13.6497
Conventional method ($\boldsymbol{G}_2(z)$)	12.9665	13.2917

TABLE II: NRR AND SNR WITH PROPOSED METHOD

	NRR [dB]	SNR_{out} [dB]
Proposed method ($\boldsymbol{G}_1(z)$)	17.2002	18.1902
Proposed method ($\boldsymbol{G}_2(z)$)	17.2004	18.1716

VII. CONCLUSION

We confirmed that our proposed method made it possible to estimate an output disturbance without the dependence of transfer characteristics if a system was a stable plant. The estimation in our method was achieved by a model whose source signals consisted of input and output disturbance, and observed signals consisted of input and output. Our experiments revealed that this method shows more accurate estimation than Miyaura's method.

REFERENCES

[1] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, John Wiley & Sons, New York, 2001.
[2] S. Ralhan, and T. A. Badgwell, "Robust output disturbance rejection for stable systems," in *Proc. the 2000 American Control Conference*, pp. 1832-1836, vol. 3, 2000.
[3] Y. Zhang, C. M. Akujuobi, W. H. Ali, C. L. Tolliver, and L. S. Shieh, "Load Disturbance resistance speed controller design for PMSM," *IEEE Trans. Ind. Electron*., vol. 53, no. 4, pp. 1198-1208, Aug. 2006.
[4] N. Murata, S. Ikeda, and A. Ziehe, "An approach to blind source separation based on temporal structure of speech signals," *Neurocomputing*, vol. 41, no. 1-4, pp. 1-24, Oct. 2001.
[5] H. Yang and S. Amari, "Adaptive On-line Learning Algorithms for Blind Separation: Maximum Entropy and Minimum Mutual Information," *Neural Computation*, vol. 9, pp. 1457-1482, 1997.
[6] S. Amari, A. Cichocki, and H. H. Yang, "A new learning algorithm for blind signal separation," *Advances in Neural Information Processing Systems*, vol. 8, MIT Press, Cambridge MA, pp. 757-763, 1997.
[7] T. Miyaura, "System Identification and Disturbance Estimation via Independent Component Analysis," M.S. thesis, Dept. Inform. Syst., Graduate School of Inform. Sci., Nara Institute of Sci. and Tech. Univ., Nara, Japan, 2002.
[8] A. Hyvärinen and E. Oja, "A fast fixed-point algorithm for independent component analysis," *Neural Computation*, vol. 9, pp. 1483-1492, 1997.
[9] J. Karhunen, E. Oja, L. Wang, R. Viario, and J. Joutsensalo, "A Class of Neural Networks for Independent Component Analysis," *IEEE Trans. Neural Network*, vol. 8, no. 3, pp. 486-503, 1997.

R. Tanaka was born in Oct. 1987 in Fukuoka, JAPAN. He received the B. E. degree in Electronics and Communications from Meiji University, Kawasaki, Japan, in 2010. He received the M. E. degree in Graduate School of Science and Technology, Meiji University, in 2012. He is currently studying toward the Dr. Eng. degree at Graduate School of Science and Technology, Meiji University. His research interests include Speech Signal Processing, Noise Reduction, and digital Control Systems

K. Kawaguchi was born in Feb. 1990 in Wakayama, JAPAN. He received the B. E. degree in Electronics and Bioinformatics from Meiji University, Kawasaki, Japan, in 2011. He is currently studying toward the M. E. degree at Graduate School of Science and Technology, Meiji University. His research interests include Dahlin Algorithm for a plant with time delay and its application to pneumatic equipment.

E. Junichi was born in Jul. 1988 in Tokyo, JAPAN. He received the B. E. degree in Electronics and Bioinformatics from Meiji University, Kawasaki, Japan, in 2011. He is currently studying toward the M. E. degree at Graduate School of Science and Technology, Meiji University. His research interests include H-infinity control and its application to DC motor.

H. Shibasaki was born in Nov. 1988 in Tokyo, JAPAN. He received the B. E. degree in Electronics and Bioinformatics from Meiji University, Kawasaki, Japan, in 2011. He is currently studying toward the M. E. degree at Graduate School of Science and Technology, Meiji University. His research interests include continuous and digital control for a plant with integrator, and an unstable plant and its application to pneumatic equipment. He is a student member of the IEEE, and the IEEJ of Japan.

Y. Hikichi was born in Oct. 1988 in Kanagawa, JAPAN. He received the B. E. degree in Electronics and Bioinformatics from Meiji University, Kawasaki, Japan, in 2011. He is currently studying toward the M. E. degree at Graduate School of Science and Technology, Meiji University. His research interests include PID control and its application to linear motor

Y. Ishida was born in Feb. 1947 in Tokyo, JAPAN. He received the B. E., M. E., and Dr. Eng. degrees in Electrical Engineering, Meiji University, Kawasaki, Japan, in 1970, 1972, and 1978, respectively. In 1975 he joined the Department of Electrical Engineering, Meiji University, as a research Assistant and became a Lecturer and an Associate Professor in 1978 and 1981, respectively. He is currently a Professor at the Department of Electronics and Bioinformatics, Meiji University. His current research interests include signal processing, speech analysis and recognition, and digital control. He is a member of the IEEE, and the IEICE of Japan.