The Detection of an Approaching Sound Source using Pulsed Neural Network

Kaname Iwasa[1], Takeshi Fujisumi[1], Mauricio Kugler[1], Susumu Kuroyanagi[1], Akira Iwata[1], Mikio Danno[2] and Masahiro Miyaji[3]

[1] Nagoya Institute of Technology, Gokiso-cho, Showa-ku, Nagoya, 466-8555, Japan
kaname@mars.elcom.nitech.ac.jp,
[2] Toyota InfoTechnology Center, Co., Ltd,
6-6-20 Akasaka, Minato-ku, Tokyo, 107-0052, Japan
[3] Toyota Motor Corporation,
1 Toyota-cho, Toyota, Aichi, 471-8572, Japan

Abstract. Current automobiles' safety systems based on video cameras and movement sensors fail when objects are out of the line of sight. This paper proposes a system based on pulsed neural networks able to detect if a sound source is approaching the sensor or moving away from it. The system, based on PN models, compares the sound level difference between consecutive instants of time in order to determine its relative movement. Moreover, the combined level difference information of all frequency channels permits to identify the type of the sound source. Experimental results show that, for three different vehicles sounds, the relative movement and the sound source type could be successfully identified.

1 Introduction

Driving safety is one of the major concerns of the automotive industry nowadays. Video cameras and movement sensors are used in order to improve the driver's perception of the environment surrounding the automobile [1][2]. These methods present good performance when detecting objects (e.g., cars, bicycles, and people) which are in line of sight of the sensor, but fail in case of obstruction or dead angles. Moreover, the use of multiple cameras or sensors for handling dead angles increases the size and cost of the safety system.

The human being, in contrast, is able to perceive people and vehicles around itself by the information provided by the auditory system [3]. If this ability could be reproduced by artificial devices, complementary safety systems for automobiles would emerge. Cause of diffraction, sound waves can contour objects and be detected even when the source is not in direct line of sight.

A possible approach for processing temporal data is the use of Pulsed Neuron (PN) models [4]. This type of neuron deals with input signals on the form of pulse trains, using an internal membrane potential as a reference for generating pulses on its output. PN models can directly deal with temporal data and can be efficiently implemented in hardware, due to its simple structure. Furthermore, high processing speeds can be achieved, as PN model based methods are usually highly parallelizable.

A sound localization system based on pulsed neural networks has already being proposed in [5] and a sound source identification system, with a corresponding implementation on FPGA, was introduced in [6]. This paper focuses specifically on the relative moving direction of a sound emitting object, and proposes a method to detect if a sound source is approaching or moving away from the sensor. The system, based on PN models, compares the sound level difference between consecutive instants of time in order to determine its relative movement. Moreover, the proposed method also identifies the type of the sound source by the use of PN model based competitive learning pulsed neural network for processing the spectral information.

2 Pulsed Neuron Model

When processing time series data (e.g., sound), it is important to consider the time relation and to have computationally inexpensive calculation procedures to enable real-time processing. For these reasons, a PN model is used in this research.

Figure 1 shows the structure of the PN model. When an input pulse $i_k(t)$ reaches the k^{th} synapse, the local membrane potential $p_k(t)$ is increased by the value of the weight w_k. The local membrane potentials decay exponentially with a time constant τ_k across time. The neuron's output $o(t)$ is given by

$$o(t) = H(I(t) - \theta) \qquad\qquad I(t) = \sum_{k=1}^{n} p_k(t) \tag{1}$$

where n is the total number of inputs, $I(t)$ is the inner potential, θ is the threshold and $H(\cdot)$ is the unit step function. The PN model also has a refractory period t_{ndti}, during which the neuron is unable to fire, indepently of the membrane potential.

3 The Proposed system

The basic structure of the proposed system is shown in Fig.2. This system consists of three main blocks, the frequency-pulse converter, the level difference extractor and the sound source classifier, from which the last two are based on PN models.

The relative movement (approaching or moving away) of the sound source is determined by the sound level variation. The system compares the signal level $x(t)$ with the level in a previous time $x(t - \Delta t)$. If $x(t) > x(t - \Delta t)$, the sound source is getting closer to the sensor, if $x(t) < x(t-\Delta t)$, it is moving away. After the level difference having been extracted, the outputs of the level difference extractors contain the spectral pattern of the input sound, which is then used for recognizing the type of the source.

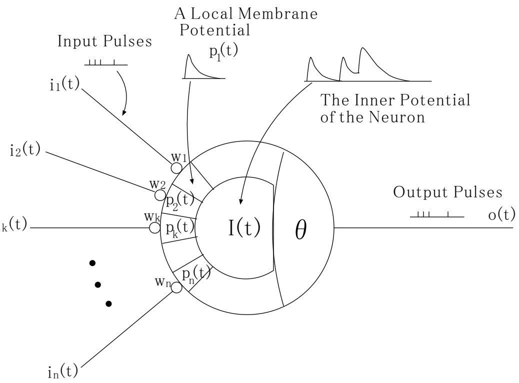


Fig. 1. Pulsed neuron model

3.1 Filtering and Frequency-Pulse Converter

Initialy, the input signal must be pre-processed and converted to a train of pulses. A bank of 4^{th} order band-pass filters decomposes the signal in 13 frequency channels equally spaced in a logarithm scale from 500 Hz to 2 kHz. Each frequency channel is modified by the non-linear function shown in Eq.(2), and the resulting signal's envelope is extracted by a 400 Hz low-pass filter. Finally, each output signal is independently converted to a pulse train, whose rate is proportional to the amplitude of the signal.

$$I(t) = \begin{cases} x(t)^{\frac{1}{3}} & x(t) \geq 0 \\ \frac{1}{4}x(t)^{\frac{1}{3}} & x(t) < 0 \end{cases} \tag{2}$$

3.2 Level Difference Extractor

Each pulse trains generated by the Frequency-Pulse converter is inputted in a Level Difference Extractor (LDE) independently. The LDE, shown in Fig. 3, is composed by two parts, the Lateral Superior Olive (LSO) model and the Level Mapping Two (LM2) model [7]. The LSO is responsible for the time difference extraction itself, while the LM2 extracts the envelope of the complex firing pattern.

Each pulse train correspondent to each frequency channel is inputted in a LSO model. The PN potential of f^{th} channel, i^{th} LSO neuron $I_{i,f}^{LSO}(t)$ is calculated as follows:

$$I_{i,f}^{LSO}(t) = p_{i,f}^{N}(t) + p_{i,f}^{B}(t) \tag{3}$$

$$p_{i,f}^{N}(t) = w_{i,f}^{N} x_f(t) + p_{i,f}^{N}(t-1)e^{-\frac{t}{\tau_{LSO}}} \tag{4}$$

$$p_{i,f}^{B}(t) = w_{i,f}^{B} x_f(t-\Delta t) + p_{i,f}^{B}(t-1)e^{-\frac{t}{\tau_{LSO}}} \tag{5}$$

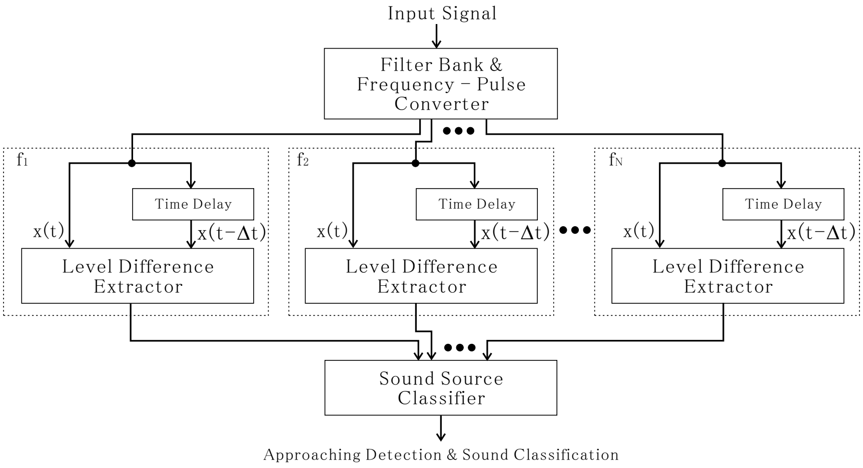

Input Signal
Filter Bank & Frequency - Pulse Converter
f1
f2
fN
Time Delay
x(t)
x(t-Δt)
Level Difference Extractor
Sound Source Classifier
Approaching Detection & Sound Classification

Fig. 2. The structure of the recognition system

where τ_{LSO} is the time constant of the LSO neuron and the weights $w^N_{i,f}$ and $w^B_{i,f}$ are defined as:

$$w^N_{i,f} = \begin{cases} 0.0 & i = 0 \\ 1.0 & i > 0 \\ -10^{\frac{i}{\gamma}} & -b < i < 0 \\ -10^{\frac{-(K-i)}{\alpha}} & i \leq -b \end{cases} \qquad w^B_{i,f} = \begin{cases} 0.0 & i = 0 \\ 1.0 & i < 0 \\ -10^{\frac{i}{\gamma}} & 0 < i < b \\ -10^{\frac{K-i}{\alpha}} & i \geq -b \ , \end{cases} \tag{6}$$

where α, γ are parameters, K is the index of the last neuron of each side of the LSO (totalizing $2K + 1$ neurons, including the central neuron) and b is the index of the last inner neuron of each side of the LSO.

As larger the signal becomes, more neurons fire on the LSO model. The LM2 stage then generates a clearer output, extracting the envelope of the firing pattern generated by the LSO. The potentials in the LM2 are calculated as follows:

$$I^{LM2}_{l,f}(t) = p^D_{l,f}(t) + p^S_{l,f}(t) \tag{7}$$

$$p^D_{l,f}(t) = m_{i,f}(t) + p^D_{l,f}(t-1)e^{-\frac{t}{\tau_{LM2}}} \tag{8}$$

$$p^S_{l,f}(t) = -m_{i,f}(t) + p^S_{l,f}(t-1)e^{-\frac{t}{\tau_{LM2}}} \tag{9}$$

where τ_{LM2} is the time constant of the LM2 neuron and $m_{i,f}(t)$ is the output of the i^{th} LSO neuron in f^{th} frequency channel.

3.3 Sound Source Classifier

The sound source classifier is based on the Competetive Learning Network using Pulsed Neurons (CONP) proposed in [5]. The basic structure of CONP is shown in Fig.4.

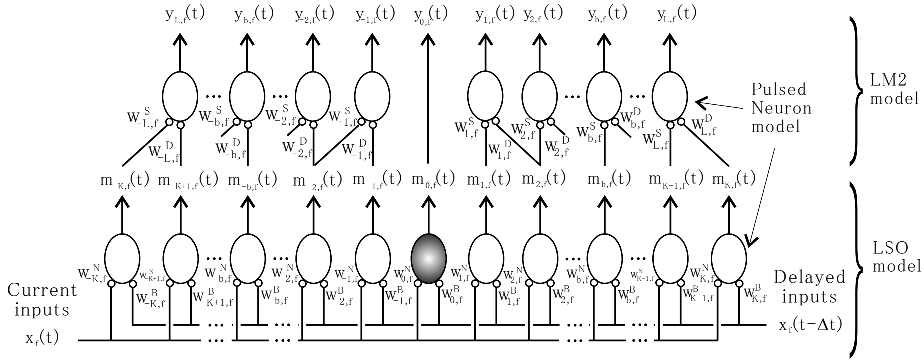


Fig. 3. Level difference extractor

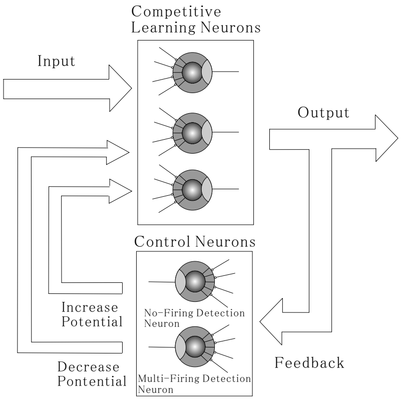

Competitive Learning Neurons
Input
Output
Control Neurons
No-Firing Detection Neuron
Multi-Firing Detection Neuron
Increase Potential
Decrease Pontential
Feedback

Fig. 4. Competitive Learning Network using Pulsed Neurons (CONP)

In the learning process of CONP, the neuron with the most similar weights to the input (winner neuron) should be chosen for learning in order to obtain a topological relation between inputs and outputs. However, in the case of two or more neurons firing, it is difficult to decide which one is the winner, as their outputs are only pulses, and not real values. In order to this, CONP has extra external units called control neurons. Based on the output of the Competitive Learning (CL) neurons, the control neurons' outputs increase or decrease the inner potential of all CL neurons, keeping the number of firing neurons equal to one. Controlling the inner potential is equivalent to controlling the threshold. Two types of control neurons are used in this work. The No-Firing Detection (NFD) neuron fires when no CL neuron fires, increasing their inner potential. Complementarily, the Multi-Firing Detection (MFD) neuron fires when two or more CL neurons fire at the same time, decreasing their inner potential.

The CL neurons are also controlled by another potential, named the input potential $p_{in}(t)$, and a gate threshold θ_{gate}. The input potential is calculated as the sum of the inputs (with unitary weights), representing the rate of the input pulse train. When $p_{in}(t) < \theta_{gate}$, the CL neurons are not updated by the control neurons and become unable to fire, as the input train has a too small potential for being responsible for an output firing. Furthermore, the input potential of each CL neuron is decreased along time by a factor β, to follow rapid changes on the inner potential and improving its adjustment.

Considering all the described adjustments on the inner potential of CONP neurons, the output equation (1) of each CL neurons becomes:

$$\begin{aligned} o(t) = H\Bigg(& \sum_{k=1}^{n} p_k(t) - \theta \\ & + p_{nfd}(t) - p_{mfd}(t) - \beta \cdot p_{in}(t) \Bigg) \end{aligned} \tag{10}$$

where $p_{nfd}(t)$ and $p_{mfd}(t)$ corresponds respectively to the potential generated by NFD and MFD neurons' outputs, $p_{in}(t)$ is the input potential and β ($0 \leq \beta \leq 1$) is a parameter.

4 Experimental Results

Three different sound sources were used on the experiments: "police car", "ambulance" and "scooter". The first two correspond to the alarm sounds of the vehicles, while the last corresponds to the engine sound of a scooter. All the signals were recorded from a static sound source. The moving sound source signals were generated by computer, with the sound intensity at each instant of time calculated as:

$$I(t) = 20 I_b \log_{10} \frac{d(t)}{d_b} \tag{11}$$

where I_b is a sound intensity in the center position, d_b and $d(t)$ are, respectively, the distance between the sensor and the sound source at center position and the distance at time t. All signal have 4.0 s of duration and the sound source is normal to the sensor at 2.0 s, as shown in Fig. 5.

4.1 Level Difference Information Extraction

The level difference information was extracted as described in section 3.2. The used parameters for the signal acquisition, preprocessing and level difference extraction are shown in Table 1.

Figure 6 shows the output of the LDE model for the "police car" signal in four distinct intervals of time. The x-axis corresponds to the index of the neurons in the LM2, representing the level difference information, and the y-axis

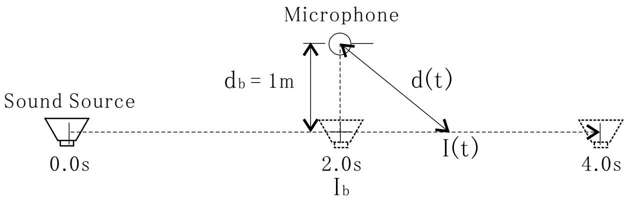


Fig. 5. Sound source movement on experiments

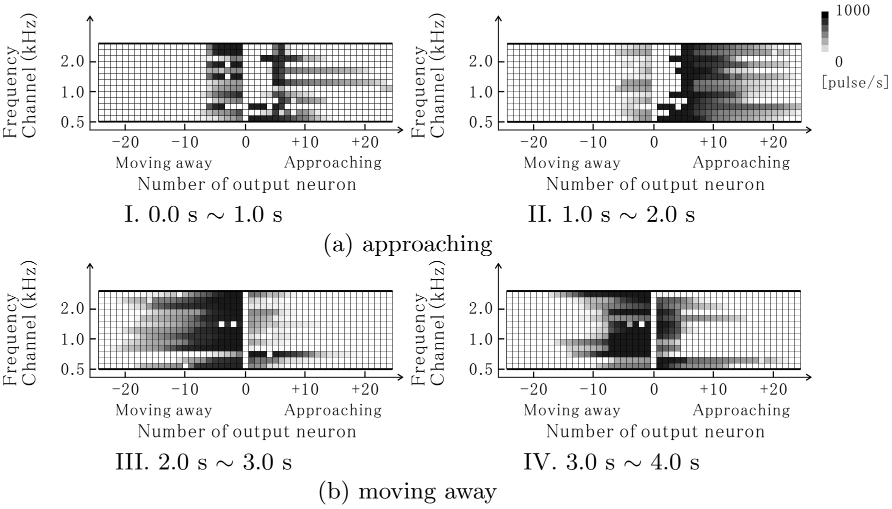
Table 1. Parameters of each module used on the experiments

Input Sound	
Sampling frequency	48 kHz
Quantization bits	16 bits
Number of frequency channels	13 channels
Delay time Δt	0.4 s
Level Difference Extractor	
Number of total LSO neurons K	51 units
Number of inner LSO neurons b	10 units
Number of output neurons L	48 units
Threshold θ_{LSO}	0.001
Threshold θ_{LM2}	0.001
Time constant τ_{LSO}	0.1 s
Time constant τ_{LM2}	35.0 μs
Parameter α	60
Parameter β	60

corresponds to the frequency channels. The gray level intensity represents the rate of the output pulse train.

The firing pattern differs significantly from each interval of time, especially when comparing the graphics of opposite relative movements. Although the LM2 could not successfully extract the envelope from the firing pattern of the signals corresponding to a sound source moving away from the sensor, the result is enough clear for distinguishing it from an approaching sound source signal.

Figure 7 shows the firing patterns of each kind of sound for the approaching (interval of 0.0 $\sim$ 2.0 s) and moving away (2.0 $\sim$ 4.0 s) cases. As different frequency components present different firing information, it is possible to classify the sound source, as described in the next section.



I. 0.0 s ∼ 1.0 s

II. 1.0 s ∼ 2.0 s

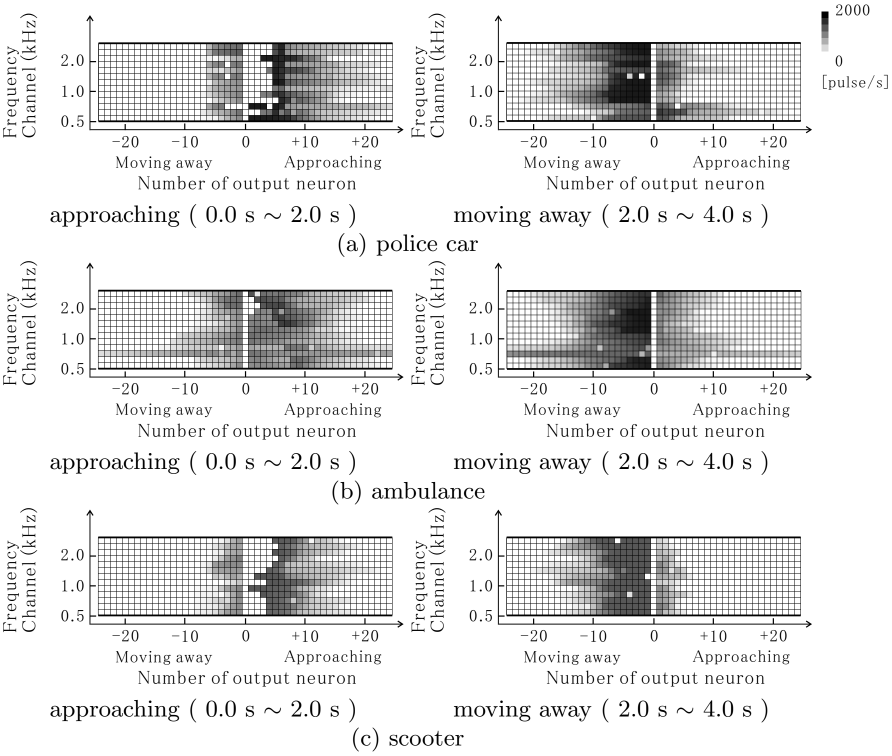
(a) approaching

III. 2.0 s ∼ 3.0 s

IV. 3.0 s ∼ 4.0 s

(b) moving away

Fig. 6. Level Difference Extractor output of the “police car” dataset



approaching (0.0 s ∼ 2.0 s)

moving away (2.0 s ∼ 4.0 s)

(a) police car

approaching (0.0 s ∼ 2.0 s)

moving away (2.0 s ∼ 4.0 s)

(b) ambulance

approaching (0.0 s ∼ 2.0 s)

moving away (2.0 s ∼ 4.0 s)

(c) scooter

Fig. 7. Comparing the output of level difference information for each dataset

Table 2. Parameters of CONP used on the experiments

Competitive learning Neuron	
Number of Inputs of CL neurons	637 units
Number of CL neurons	30 units
Threshold θ	$1.0{\times}10^{-4}$
Gating threshold θ_{gate}	150.0
Rate for input pulse frequency β	0.0629
Time constant τ_p	0.1 s
Refractory period t_{ndti}	10 ms
Learning coefficient α	$2.0{\times}10^{-8}$
Learning iterations	1000
No-Firing Detection Neuron	
Time constant τ_{NFD}	0.5 ms
Threshold θ_{NFD}	$-1.0{\times}10^{-3}$
Connection weight to each CL neurons	16.0
Multi-Firing Detection Neuron	
Time constant τ_{MFD}	1.0 ms
Threshold θ_{MFD}	2.0
Connection weight from each CL neurons	16.0

4.2 Sound Source Classification

The firing information patterns provided by all the level difference extractors are recognized by the CONP model described in section 3.3. The CONP model was trained according to the parameters shown in Table 2. Table 3 shows the accuracy of the CONP model for each dataset. The recognition rate is defined as the ratio between the number of neuron's firing corresponding to the correct vehicle and relative movement and the total number of firings. The correct sound source and relative direction could be recognized with an average accuracy of 85.8%.

The results of the "scooter" dataset present a better recognition rate than the "police car" and "ambulance" datasets. The reason for this is that the sound signal of the "scooter" dataset is constant over time, in opposite to the alarm sounds of the other two vehicles, which actually correspond to two different and alternated sounds. Thus, the CONP model can be more efficiently trained with the "scooter" data than the others, which would require more data in order to obtain a comparable accuracy.

5 Conclusions

This paper proposes a system for detecting the approaching and classifying a sound source using pulsed neural networks. The system extracts the level difference information from pulse trains corresponding to several frequency bands. The firing pattern is then classified by a CONP model, which identifies the type and recognizes the relative movement of the sound source.

Table 3. Results of sound recognition
(A = Approaching , M = moving away)

		Recognition Rate[%]					
		police car		ambulance		scooter	
Input Sound		A	M	A	M	A	M
police	A	70.6	6.8	2.4	7.3	12.9	0.0
	M	6.8	88.3	0.0	4.9	0.0	0.0
ambulance	A	1.1	4.2	82.8	9.9	2.0	0.0
	M	3.8	0.2	7.3	86.3	0.0	2.4
scooter	A	0.0	0.0	5.7	0.0	94.3	0.0
	M	0.0	1.9	0.3	5.4	0.0	92.4

The experimental results confirmed that the PN model based level difference extractor can successfully detect the relative movement (approaching or moving away) of a sound source. By using the firing pattern provided by the LDE, the sound source type and relative movement could be correctly classified with a average accuracy of 85.8%.

Future works include the detection of a sound source position (its distance from the sensor) and the combination of the proposed system with a sound localization method. The hardware implementation of the proposed systems using an FPGA device is also in progress.

Acknowledgment

This research is supported in part by a grant from the Hori Information Science Promotion Foundation.

References

1. Surendra G., Osama M.,Robert F.K.M, Nikolaos P.P. : “Detection and Classification of Vehicles”, IEEE Trans. ITS, vol.3, No.1, pp.37-47, 2002.
2. Chieh-Chi W. Thorpe C. Thrun S. : “Online simultaneous localization and mapping with detection and tracking of moving objects: theory and results from a ground vehicle in crowded urban areas”. Proceedings of ICRA 2003, pp.842-849, 2003.
3. Pickles J.O.: “An Introduction to the Physiology of Hearing” , ACADEMIC PRESS, 1988.
4. Maass W., and Bishop C.M., : “Pulsed Neural Networks”, MIT Press , 1998.
5. Kuroyanagi, S. , Iwata, A. : “A Competitive Learning Pulsed Neural Network for Temporal Signals”, Proceedings of ICONIP 2002, pp.348-352, 2002.
6. Iwasa, K., Kuroyanagi, S. , Iwata, A. : “A Sound Localization and Recognition System using Pulsed Neural Networks on FPGA”, to appear in : Proceeding of International Joint Conference of Neural Networks 2007, August, 2007.
7. Kuroyanagi, S. , Iwata, A. : “Auditory Pulse Neural Network Model to Extract the Inter-Aural Time and Level Difference for Sound Localization”, IEICE Trans. Information and Systems, E77-D, No.4, pp.466-474, 1994.
8. Kuroyanagi, S. , Iwata, A. : “Auditory Pulse Neural Network Model for Sound Localization -Mapping of the ITD and ILD-”, IEICE J78-D2, No.2, pp.267-276, 1996(in Japanese).