PAPER

An Approach for Sound Source Localization by Complex-Valued Neural Network

Hirofumi TSUZUKI[†a)], ***Student Member***, **Mauricio KUGLER**[†b)], ***Nonmember***, **Susumu KUROYANAGI**[†c)], ***and*** **Akira IWATA**[†d)], ***Members***

SUMMARY This paper presents a Complex-Valued Neural Network-based sound localization method. The proposed approach uses two microphones to localize sound sources in the whole horizontal plane. The method uses time delay and amplitude difference to generate a set of features which are then classified by a Complex-Valued Multi-Layer Perceptron. The advantage of using complex values is that the amplitude information can naturally masks the phase information. The proposed method is analyzed experimentally with regard to the spectral characteristics of the target sounds and its tolerance to noise. The obtained results emphasize and confirm the advantages of using Complex-Valued Neural Networks for the sound localization problem in comparison to the traditional Real-Valued Neural Network model.

key words: *sound source direction, complex-valued neural networks, binaural sound localization, amplitude-phase function*

1. Introduction

Sound is a very important factor in life. People can distinguish different sounds from different directions, which makes it possible to understand their surroundings and detect dangerous situations. Therefore, various studies have been conducted in order to realize this functionality.

Among these studies, Multiple Signal Classification (MUSIC) [1] is widely used; it uses a microphone array to estimate the direction of arrival of given sound. This method is based on a correlation matrix obtained from several signals recorded by the microphone array. Estimation of Signal Parameters via Rotation Invariance Techniques (ESPRIT) [2] is another method that uses microphone arrays. While this method is also based on a correlation matrix, it has the advantage of not requiring any calibration for the microphone array. The aforementioned methods use several microphones in order to localize a sound source. On the other hand, a different approach for sound source localization is to use only two microphones since animals can localize sound sources with two ears [3], [4]. These methods estimate the direction of a sound source using Maximum-Likelihood Estimation (MLE) or a variation of MUSIC.

Furthermore, neural networks are also widely used for sound source localization. Among these, the most popular are the Radial Basis Function (RBF) networks used with microphone arrays [5], [6]. Neural networks have also been applied in a two-microphone configuration [7].

Complex-valued neural networks have been used in combination with microphone arrays for sound localization [8], however, the potential for their use in a two-microphone configuration is yet to be explored.

The advantage of using complex-valued neural networks which can treat complex numbers in their original representation is particularly important in the sound localization domain. For instance, when using real numbers, the sound phase information has to be represented in a bounded interval of $[-\pi$ and $\pi]$. However, by using complex numbers, the sound phase information can be represented correctly without these artificial edges.

Based on the aforementioned suitability, this paper proposes a sound localization method that uses complex-valued neural networks in a two-microphone configuration.

2. Background

2.1 Sound Localization

The most basic information used for determining the direction of a sound source is the difference of sound pressure and time of arrival between the two microphones. Because the time delay information between two microphones for a given direction takes the shape of a hyperbola [9], the localization of the sound source is limited to a half-plane. Thus, for localization across the whole plane, the sound pressure difference information is also necessary (Fig. 1).

This information is then classified using a Complex-Valued Multi-Layer Perceptron (CV-MLP) neural network.

2.2 Complex-Valued MLP

Complex-valued neural networks [10], [11] behave like standard neural networks but allow the calculation using complex numbers domain. While the basic construction of a CV-MLP is similar to that of a real-valued MLP (RV-MLP) [12], few key differences exist with regard to the activation function and the learning method. Although the sigmoid and Gaussian functions tend to be the most commonly adopted activation functions for RV-MLP, they are inappropriate for complex number inputs as they may result

Manuscript received January 18, 2013.
Manuscript revised April 24, 2013.

[†]The authors are with the Department of Scientific and Engineering Simulation, Nagoya Institute of Technology, Nagoya-shi, 466–8555 Japan.
a) E-mail: hrfm.tsuzuki@gmail.com
b) E-mail: mauricio@kugler.com
c) E-mail: bw@nitech.ac.jp
d) E-mail: iwata@nitech.ac.jp

DOI: 10.1587/transinf.E96.D.2257

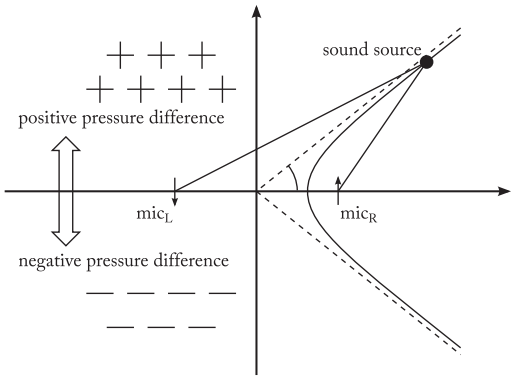


Fig. 1 Time delay and pressure difference information in the two-microphone configuration.

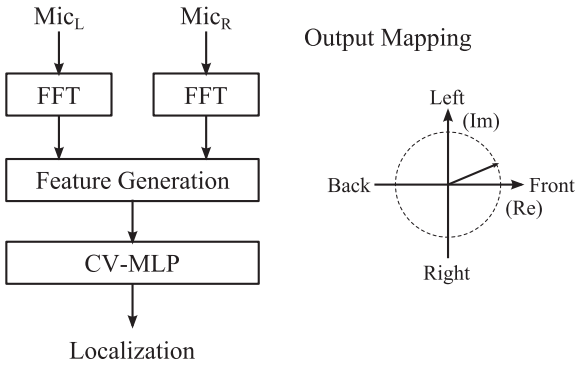


Fig. 2 Basic structure of the proposed method and output mapping of the complex values in the horizontal plane.

in infinite values. Therefore, the following equations will be adopted as the activation functions for this CV-MLP.

$$f(u) = \tanh(u_{\text{real}}) + i\tanh(u_{\text{imag}}) \tag{1}$$

$$f(u) = \tanh(|u|)e^{i\arg u} \tag{2}$$

where u is the summation of the weighted input values. In the adopted model of CV-MLP, the amplitude-phase function in Eq. (2) is used. Thus, the learning proceeds as follows:

$$\left|w_{kj}^{\text{new}}\right| = \left|w_{kj}^{\text{old}}\right| - \alpha\frac{\partial E}{\partial|w_{kj}|} \tag{3}$$

$$\arg w_{kj}^{\text{new}} = \arg w_{kj}^{\text{old}} - \alpha\frac{1}{|w_{kj}|}\frac{\partial E}{\partial(\arg w_{kj})} \tag{4}$$

$$\begin{aligned}\frac{\partial E}{\partial|w_{kj}|} = {} & \left(1-|z_k|^2\right)\left(|z_k| - |\hat{z}_k|\cos(\arg z_k - \arg\hat{z}_k)\right)|z_j| \\ & \cdot\cos(\arg z_k - \arg z_j - \arg w_{kj}) \\ & - |z_k||\hat{z}_k|\sin(\arg z_k - \arg\hat{z}_k)\frac{|z_j|}{\tanh^{-1}|z_k|} \\ & \cdot\sin(\arg z_k - \arg z_j - \arg w_{kj})\end{aligned} \tag{5}$$

$$\begin{aligned}\frac{1}{|w_{kj}|}\frac{\partial E}{\partial(\arg w_{kj})} = {} & \left(1-|z_k|^2\right)\left(|z_k| - |\hat{z}_k|\cos(\arg z_k - \arg\hat{z}_k)\right)|z_j| \\ & \cdot\sin(\arg z_k - \arg z_j - \arg w_{kj}) \\ & - |z_k||\hat{z}_k|\sin(\arg z_k - \arg\hat{z}_k)\frac{|z_j|}{\tanh^{-1}|z_k|} \\ & \cdot\cos(\arg z_k - \arg z_j - \arg w_{kj})\end{aligned} \tag{6}$$

where $j = 1, \cdots, J$ and $k = 1, \cdots, K$ are number of hidden layer neuron and output layer neuron, w_{kj} is a weight of the synapse between neuron j and k, E is the error between the desired output $\hat{z}$ and the the output value z. The weights of the output layer neurons are updated in Eq. (3) and (4), and the derivatives are calculated in Eq. (5) and (6). In the case of a CV-MLP, it is not sufficient enough to solely back-propagate the error, as in an RV-MLP. It is also necessary to back-propagate the desired outputs in order to update the weights of the neurons in the hidden layer. The desired outputs for the hidden layer are calculated as follows:

$$\hat{z}_j^{\text{H}} = \overline{f\left(\sum_k \overline{\hat{z}_k^{\text{O}}}\cdot w_{kj}\right)} \tag{7}$$

where H and O mean hidden layer and output layer, $\overline{z}$ notation means a complex conjugate and f is an activation function (2).

3. Proposed Method

The proposed method consists of three main blocks: the signal pre-processing block, the feature generation block, and the CV-MLP. The pre-processing block calculates the frequency-phase spectrum in the complex form using fast Fourier transform (FFT). The feature generation block, explained in detail in the following section, transforms the FFT output for efficient use with the CV-MLP. Finally, the results of the CV-MLP are associated with sound source directions. The main structure is shown in Fig. 2.

3.1 Feature Generation

The phase and amplitude information generated by the pre-processing block for the left and right channels are calculated as follows:

$$\begin{aligned}\mathbf{r}^L e^{\boldsymbol{\theta}^L i} &= \left[r_0^L e^{\theta_0^L i}, \cdots, r_{\frac{N}{2}-1}^L e^{\theta_{\frac{N}{2}-1}^L i}\right] \\ \mathbf{r}^R e^{\boldsymbol{\theta}^R i} &= \left[r_0^R e^{\theta_0^R i}, \cdots, r_{\frac{N}{2}-1}^R e^{\theta_{\frac{N}{2}-1}^R i}\right]\end{aligned} \tag{8}$$

where $\mathbf{r}^L e^{\boldsymbol{\theta}^L i}$ and $\mathbf{r}^R e^{\boldsymbol{\theta}^R i}$ are vectors of complex numbers generated by the FFT, representing amplitude and phase, and N is the FFT's window length. Sound level is usually represented in a logarithmic scale which allows the evaluation of the sound pressure difference, regardless of the absolute sound level. As the sound level difference is represented by the ratio of two sounds, a logarithmic scale is necessary.

If a logarithmic scale is not used, the sound pressure difference is affected by the absolute sound level. However, a logarithmic scale would include negative values of amplitude. As a replacement, the cubic root, which behaves similarly as the logarithmic scale, was used.

In order to obtain the time delay, the phase difference is calculated as follows:

$$\begin{cases} \mathbf{s}^L e^{(\boldsymbol{\theta}^R - \boldsymbol{\theta}^L)i} \\ \mathbf{s}^R e^{(\boldsymbol{\theta}^L - \boldsymbol{\theta}^R)i} \end{cases} \tag{9}$$

where $\mathbf{s}^L$ and $\mathbf{s}^R$ are respectively:

$$\mathbf{s}^L = \sqrt[3]{\mathbf{r}^L}, \quad \mathbf{s}^R = \sqrt[3]{\mathbf{r}^R} \tag{10}$$

The phase difference is linearly proportional to the frequency. If each phase difference is rotated by a correct angle, all phase differences will become 0 radians. Otherwise, the phase differences will take different values. In case of correct rotation angles, the summation of values in Eq. (9) will result in a large amplitude. In the other case, the summation will result in a small amplitude. Furthermore, frequencies with small amplitudes do not affect the summation significantly. In other words, the amplitude behaves as a mask of phase difference.

The complete feature vector is then given by $\mathbf{F} = [F_0, \cdots, F_{D-1}]$. The number of features D corresponds to the number of subdivisions of the interval $[-\Phi, \Phi]$, where the optimal value of D is determined experimentally. Each d^{th} element of the vector $\mathbf{F}$ is calculated by subtracting the aforementioned summation of the right signal from the left one, as follows:

$$F_d = \sum_{k=0}^{\frac{N}{2}-1} s_k^L e^{(\theta_k^R - \theta_k^L + \frac{k}{\frac{N}{2}-1}\phi_d)i} - \sum_{k=0}^{\frac{N}{2}-1} s_k^R e^{(\theta_k^L - \theta_k^R - \frac{k}{\frac{N}{2}-1}\phi_d)i} \tag{11}$$

where s_k^L and s_k^R are, respectively, the k^{th} elements of $\mathbf{s}^R$ and $\mathbf{s}^R$ given by Eq. (9), and the rotation angle ϕ_d is the d^{th} element of $[-\Phi, \Phi]$. The subtraction sign indicates which microphone has the larger amplitude, and thus, sounds coming from the front and back half-planes can be distinguished.

Figure 3 shows examples of feature vectors generated from white noise coming from four different directions. It is clear that each different direction corresponds to a different value of ϕ. Figure 4 (a) and (b) are enlarged portions of Fig. 3 (c) and (d), respectively.

3.2 Localization

As stated before, the localization is performed by a CV-MLP. In the proposed method, an amplitude-phase activation function is used as follows:

$$f(z) = \tanh(|z|)e^{i \arg z} \tag{12}$$

The CV-MLP has three layers: an input layer, a hidden layer and an output layer. The number of neurons on the input layer is the same as the dimension of the feature vector. The final output phase corresponds to the sound source direction. Hence, a complex number of amplitude 1.0 is used as the desired output. For instance, when a sound source direction of 45 degrees is inputted, $e^{\frac{\pi}{4}i}$ is used as the desired output.

The order of complexity of the proposed method can be obtained by analysing each of its blocks independently. The FFT-based signal preprocessing block has complexity $O(N \log N)$, while the feature generation block has complexity $O(ND)$. The CV-MLP has complexity $O(DH)$, where H is the number of hidden neurons. By contrast, the widely used correlation method has complexity $O(N^2)$. In the proposed method, the feature vector dimension D is usually much smaller than the window length N. Therefore, the computation cost for the proposed feature generation method is considerably smaller than the standard correlation method.

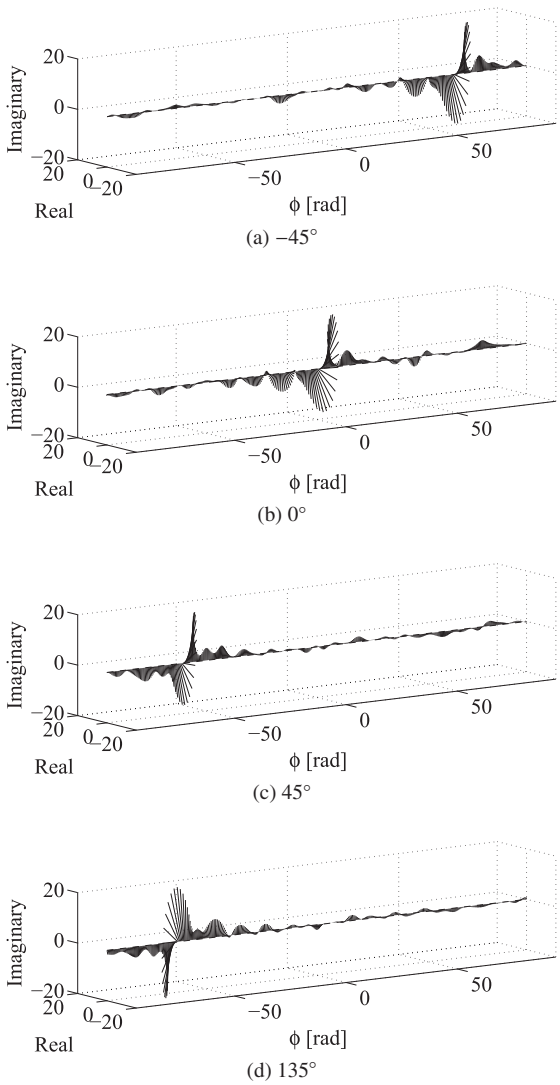
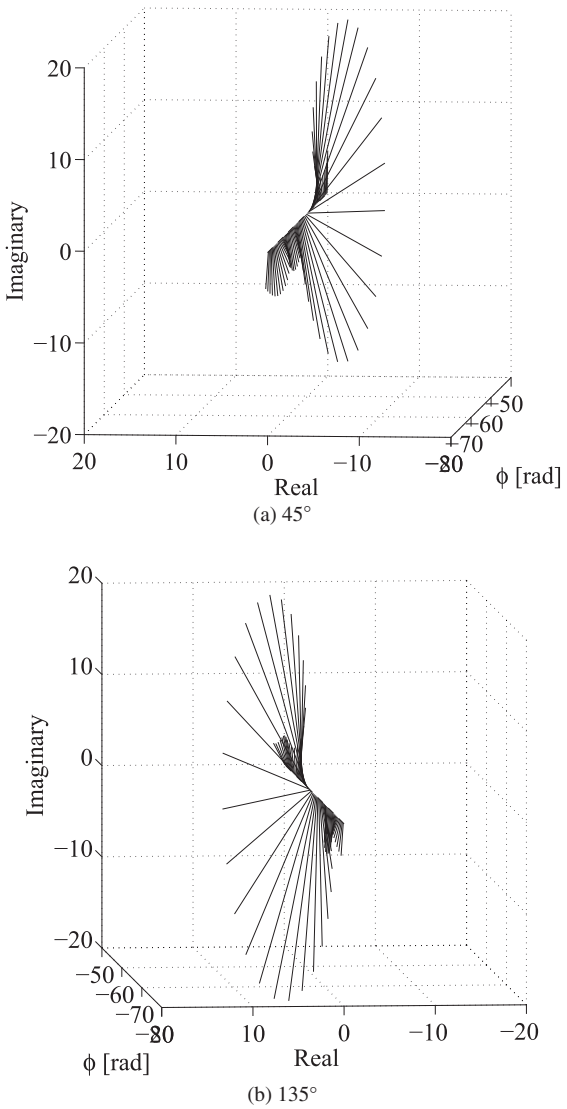


Fig. 3 Extracted features for white noise localized in different directions.



(a) 45°

(b) 135°

Fig. 4 Detailed view of the extracted features (white noise) for two different directions.

Table 1 Experimental conditions.

Recording	
microphones distance (m): L_{mic}	0.20
speaker distance (m): L_{spk}	2.00
sampling frequency (Hz): f_{s}	48000
Feature generation	
window length: N	16384
rotation (rad): Φ	30π
dimension: D	400
CV-MLP	
hidden layer dimension H	20
learning iteration	200

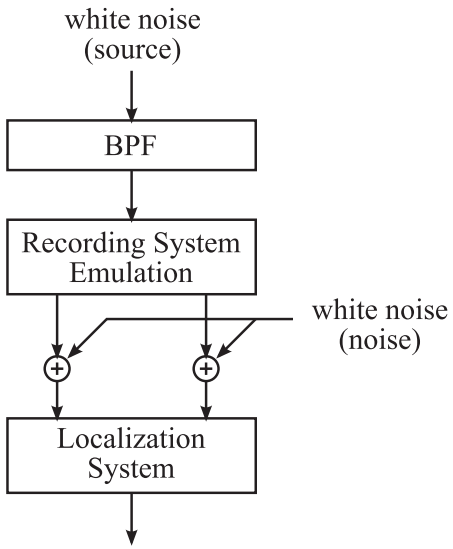


Fig. 5 Emulation system for band-limited signals with added noise.

4. Experiments

Initially, artificial sound sources were used for evaluation of the proposed method. The same experiments were conducted for both, CV-MLP and RV-MLP, to verify the advantages of using a CV-MLP.

Then, experiments were conducted using real sounds. When recording the real sounds, MX184 (SHURE) and MSP7 STUDIO (YAMAHA) were used for the recording system.

Table 1 shows experimental conditions. The speaker distance L_{spk} represents the distance from the speaker to the midpoint between the microphones. The rotation Φ must exceed the maximum phase difference of the sound from the two microphones. Therefore, the rotation Φ is set to a value greater than the phase difference of the Nyquist frequency f_{nyq}, in the case where the sound source is arranged on a straight line and two microphones. This value is calculated as follow:

$$\Phi \geq \frac{2\pi L_{\mathrm{mic}} f_{\mathrm{nyq}}}{V}\Big|_{V=342} \simeq 28\pi \tag{13}$$

where V is a velocity of sound.

In the experiment, the proposed system localizes sound source to 8 directions.

4.1 Artificial Sounds

The experiment evaluates the CV-MLP and the RV-MLP with regard to several frequency ranges and signal-to-noise ratios (SNR).

Figure 5 shows the experiment flow. The frequency range of the white noise sound source is limited by a band pass filter (BPF). This limited noise is then inputted into the emulation system which introduces a time delay and a sound pressure gain. The delay and the pressure gain were obtained using the distance between the microphone and the speaker and the microphone's polar pattern. The delay T and the pressure gain A are obtained as follows:

Table 2 Test sounds.

cut frequency ratio	$1/20, 2/20, \cdots, 20/20$
S/N ratio (dB)	-6, 0, 6, 12, 18, 24, ∞

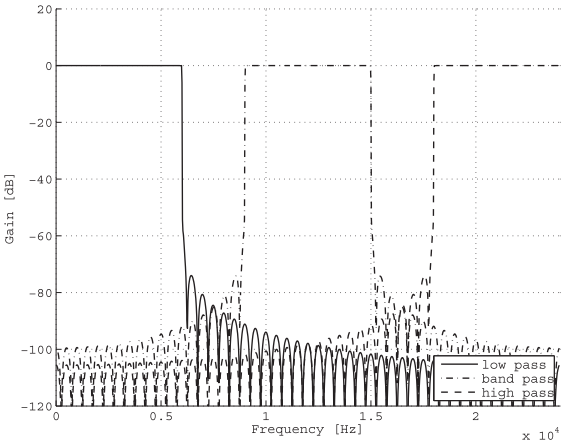


Fig. 6 Filter's frequency response (band ratio: 5/20).

$$T = \left[\frac{\left|\overrightarrow{MS}\right| fs}{V} \right] \tag{14}$$

$$A = g(\arg \overrightarrow{MS}) - 20 \log_{10} \left|\overrightarrow{MS}\right| \tag{15}$$

where $\overrightarrow{MS}$ is a vector from the microphone to the sound source and $g(\cdot)$ is the microphone's polar pattern function. The emulated sound $x'(t)$ is generated from the sound source x as follows:

$$x'(t) = Ax(t - T) \tag{16}$$

After emulation, uncorrelated white noise is added to the emulated sounds as environmental noise.

Table 2 shows the parameters used to generate the artificial sounds. The cut frequency ratio is the ratio between the filter pass-band range and the Nyquist frequency. For example, for a sampling frequency of 48000 Hz, a cut frequency ratio of 3/20 means 3600 Hz bandwidth with the 24000 Hz Nyquist frequency. Actually, three types of filters were used: low-pass, band-pass and high-pass. Figure 6 shows an example of each of the filters' frequency response. The center frequency of the band-pass filter is half of the Nyquist frequency. The SNR is measured only within the filter pass range. The signal level is the average sound level of the white noise emulated by the recording system, while the noise level is the sound level of the white noise added after the emulation system. Thus, the obtained values of SNR are larger than they would have been if calculated across the whole spectrum. The reason for this is that the SNR at the stop range of the filters has a negative infinite value, since only noise is present.

At first, for each combination of parameters of Table 2, an input sound signal was generated. Each signal was used to train and test the CV-MLP independently. This process was repeated for each filter type. The obtained results are shown in Fig. 7.

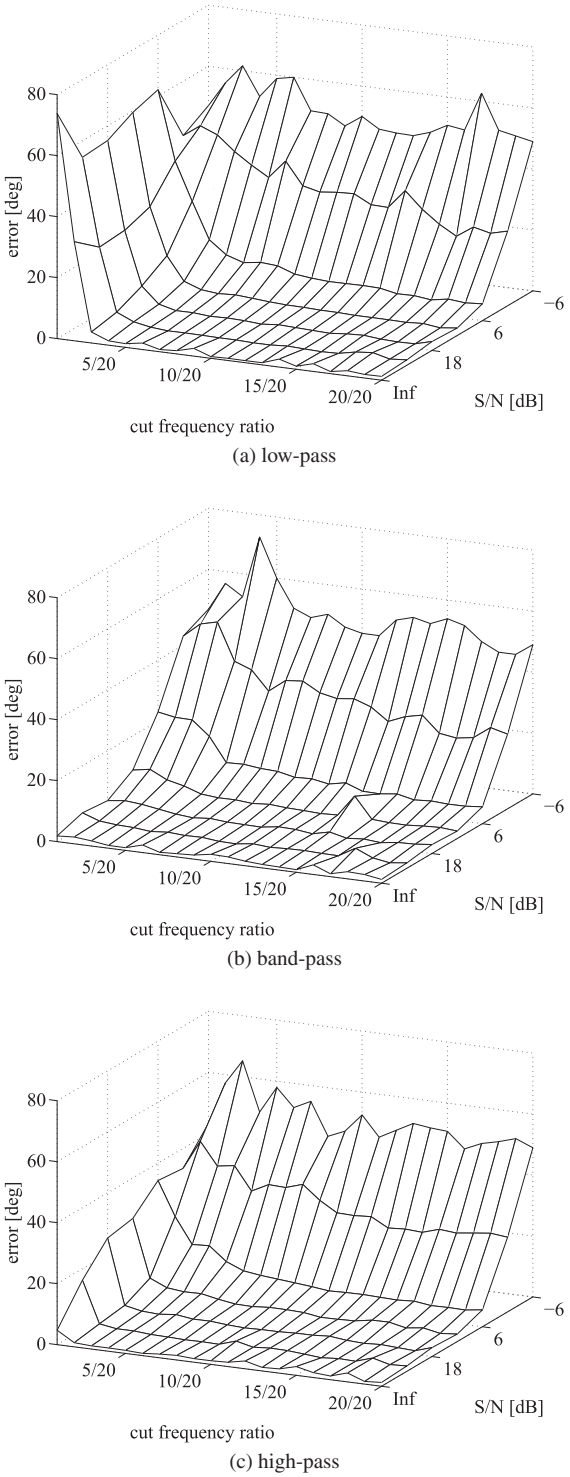
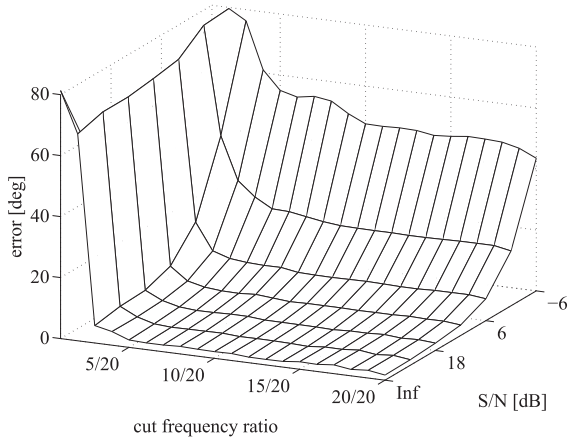
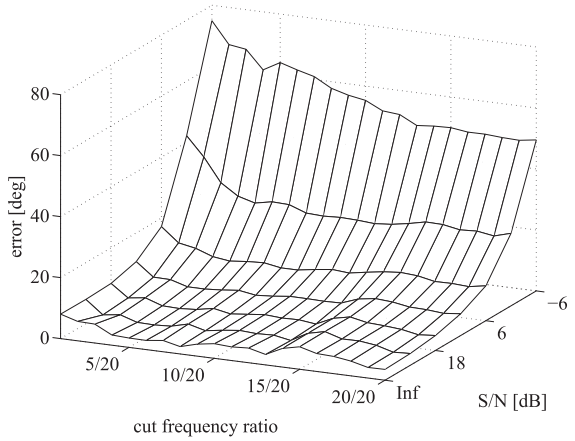


Fig. 7 Localization error for a single training set.

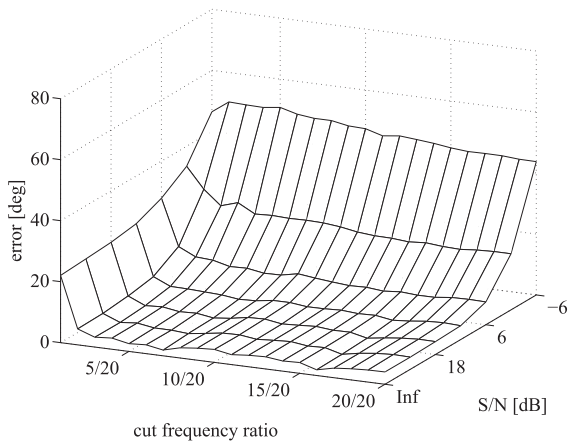
Then, all signals for which the error was below 5 degrees were combined and used to retrain the CV-MLP. Unlike the training, the test set also included the signals for which the error was larger. The results are shown in Fig. 8. Finally, the training sets for all filter types were combined and used to re-train the CV-MLP. The complete set of



(a) low-pass



(b) band-pass



(c) high-pass

Fig. 8 Localization error for the combined training sets.

sounds was then used to test the CV-MLP, of which the results are shown in Fig. 9.

From the experimental results, it is clear that the value of SNR needs to exceed 6 dB in order to achieve a successful sound localization. For lower SNR values, it becomes difficult to obtain consistent time delays. Moreover, narrow frequency ranges also make it difficult to obtain the correct

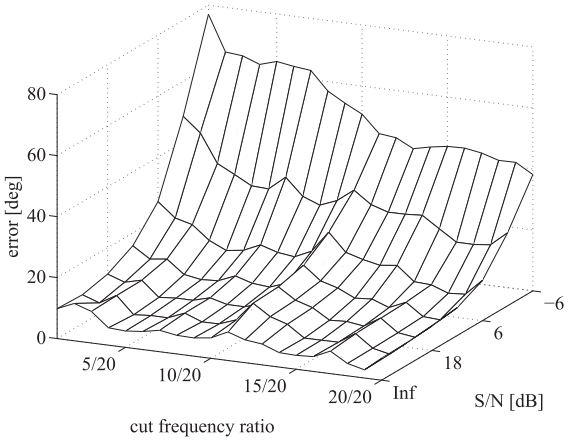


Fig. 9 Localization error for the combination of all filters.

time delay, since several values of delay can be estimated from a single phase difference. Furthermore, when compared to high frequencies, despite the same time delay, much smaller phase differences are obtained in the case of low frequencies, making the estimation of the time delay less accurate. The results indicate that low frequency sounds require a bandwidth ratio of at least 3/20, while high frequency sounds only require a minimal bandwidth ratio of 1/20.

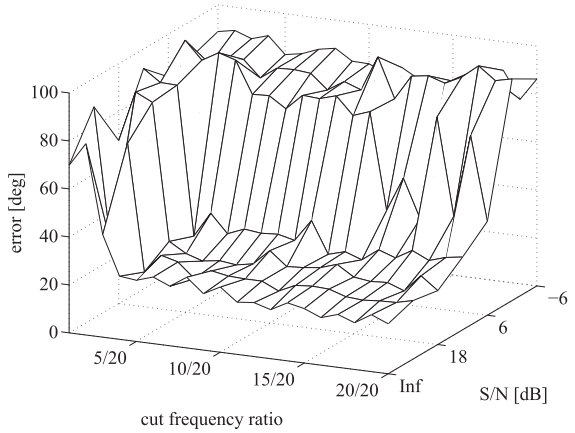
4.2 Comparison with Real-Valued MLP

The CV-MLP's performance was compared to that of the RV-MLP. The same data was used for training and testing both models. The number of neurons in the hidden and output layers is the same for both models. The input layer of the RV-MLP has double the dimension of that of the CV-MLP, as the complex values of the features have to be divided into two real values (amplitude and phase). In the RV-MLP the sigmoid function was used as the activation function.

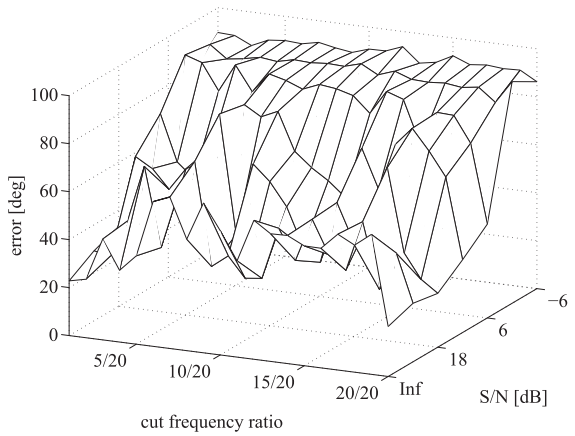
Figure 10 shows the accuracy of the RV-MLP. Clearly, those results are significantly inferior to the results of the CV-MLP shown in Fig. 7. For the purpose of our method, it is necessary to consider a combination of amplitude and phase as the features. The inferior results of the RV-MLP can thus be explained by the fact that it can only consider amplitude and phase independently. Hence, it is sound to conclude that complex-valued neural networks are the more appropriate approach for the sound localization problem.

4.3 Real Sounds

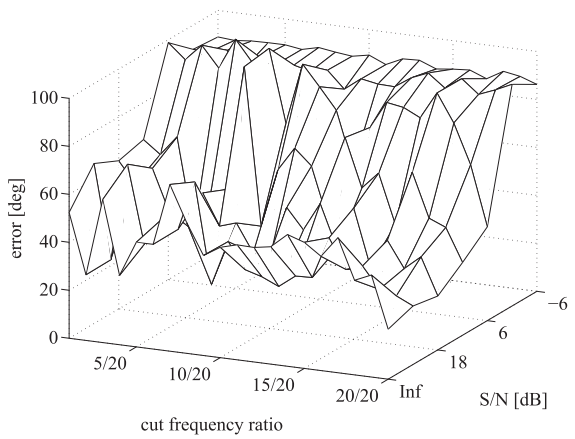
The proposed method was also tested with 21 environmental sounds listed in Table 3, in which the bold names correspond to sounds with two different variations. The sounds were played through a speaker and recorded using two cardioid microphones arranged according to the diagram in Fig. 11. The two microphones are fixed to the ends of a bar facing opposite directions as shown Fig. 11 (a). When using a cardioid microphone, signals coming from the back of the microphone are inverted due to the sensor's characteristics. In



(a) low-pass



(b) band-pass



(c) high-pass

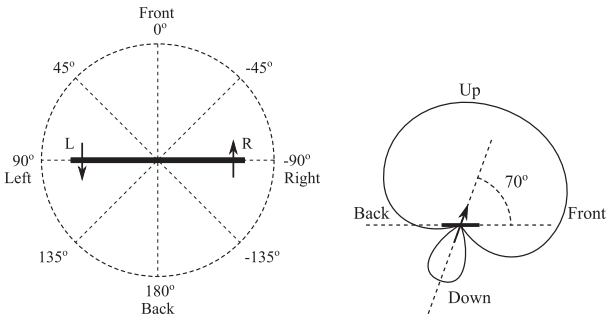
Fig. 10 Localization error for RV-MLP.

this case, the time difference cannot be calculated properly. To avoid this, the microphones are tilted 70 degrees upwards as shown Fig. 11 (b). By doing this, the signals from both front and back present the same phase, yet differ in amplitude. Figure 11 (a) defines the sound direction angles referenced hitherto.

Figure 12 shows localization accuracy for the proposed

Table 3 Sound types.

pure tone *(250Hz, 500Hz, 1kHz, 2kHz, 4kHz)*
alarm
siren (**ambulance**, fire truck, **police car**)
interphone
kettle
mobile phone
phone
voice
white noise



(a) top view (b) side view

Fig. 11 Microphone configuration used to record real sounds.

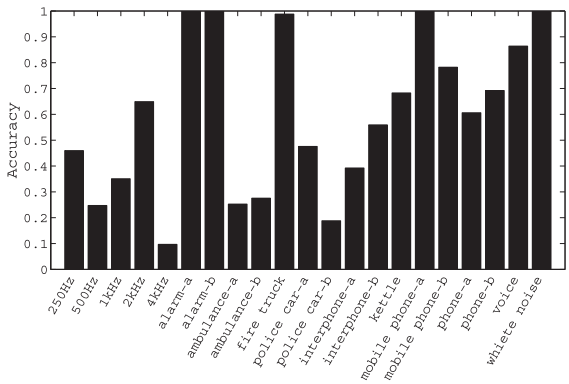
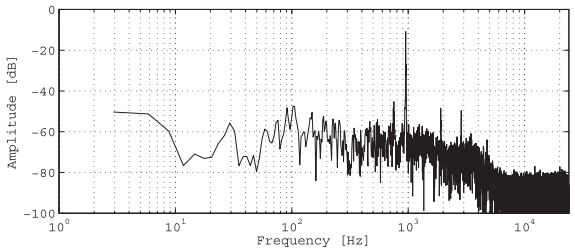
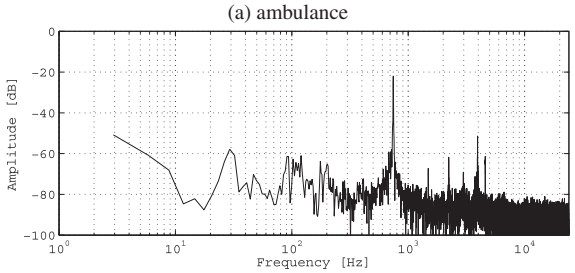


Fig. 12 Localization accuracy of real sounds.

method with different sounds. For the purpose of this experiment, directions are divided into eight sectors of 45 degrees each. Therefore, an absolute error of up to 22.5 degrees is accepted as a correct answer. From Fig. 12, the accuracy of mobile (a) and white noise was 100%. The accuracy of fire truck was also very high. On the other hand, the accuracy of pure tones was considerably low. That can be explained by the fact that pure tones contain only one single frequency from which only one phase difference can be extracted. However, in order to calculate a time delay, the phase information of at least two different frequencies is needed. Nevertheless, some of the pure tones present non-zero accuracy. As the propose method treats the localization problem as a pattern recognition problem, some of the data can be classified to the correct directions due to small differ-



(a) ambulance



(b) interphone

Fig. 13 Power spectrum of low accuracy sounds.

ences among the feature vectors.

Figure 13 shows the power spectrum of the siren (ambulance) and interphone. Each sound contains only one frequency component of around 950 Hz and 750 Hz respectively. Thus, because the restricted spectrum of the sound approaches a pure tone, correct time delay information could not be obtained. On the other hand, sounds with slightly more varied spectral components produce fairly accurate results.

5. Conclusions

This paper presented a novel approach for sound localization based on complex-valued neural networks. This method uses time delay, phase delay and sound pressure difference information, extracted from the sound signals acquired by two microphones, in order to generate the features. The features are then classified by a CV-MLP. In the experiments, the proposed method was used to localize sound sources in eight different directions distributed along both the front and the rear planes.

Artificial sounds with at least 6 dB SNR were successfully localized by the system. In the case of low-frequency sounds, a bandwidth ratio of at least 3/20 is needed for successful localization. This behavior was confirmed by experiments using real sounds in which the sounds with varied spectral components presented a higher accuracy than the ones with narrow spectra. This is explained by the fact that multiple frequency components are required for extracting meaningful time delay information. Moreover, the CV-MLP model consistently outperformed its RV-MLP counterpart. This is due to the real-valued neural networks' treatment of the phase and the amplitude as independent information.

Future research aims at improving the resolution of the system from the current 45 degrees. In addition, while the proposed method deals with sound source localization as a classification problem, it is possible to treat it as an estimation problem since the CV-MLP can naturally generate analog outputs.

References

[1] R.O. Schmidt, "Multiple emitter location and signal parameter estimation," IEEE Trans. Antennas Propag., vol.34, no.3, pp.276–280, March 1986.
[2] R. Roy and T. Kailath, "Esprit–estimation of signal parameters via rotational invariance techniques," IEEE Trans. Acoust. Speech Signal Process., vol.37, no.7, pp.984–995, 1989.
[3] N. Ono, S. Fukamachi, T. Nihsimoto, and S. Sagayama, "Sound source localization for whole direction by asymmetrically arrayed 2ch microphones on sphere," IEICE Technical Report, vol.106, no.614, pp.25–30, March 2007.
[4] Y. Nagata, S. Iwasaki, T. Hariyama, H. Horiguchi, T. Fujioka, and M. Abe, "Step by step doa estimation of multiple sound sources based on music with two-channel input," IEICE Trans. Fundamentals (Japanese Edition), vol.J92-A, no.11, pp.864–873, Nov. 2009.
[5] A.H.E. Zooghby, C.G. Christodoulou, and M. Georgiopoulos, "A neural network-based smart antenna for multiple source tracking," IEEE Trans. Antennas Propag., vol.48, no.5, pp.768–776, May 2000.
[6] S. Vigneshwaran, N. Sundararajan, and P. Saratchandran, "Direction of arrival (doa) estimation under array sensor failures using a minimal resource allocation neural network," IEEE Trans. Antennas Propag., vol.55, no.2, pp.334–343, Feb. 2007.
[7] M. Kugler, T. Hishida, S. Kuroyanagi, and A. Iwata, "A novel approach for hardware based sound localization," ICANN (1), ed. K.I. Diamantaras, W. Duch, and L.S. Iliadis, Lecture Notes in Computer Science, vol.6352, pp.490–499, 2010.
[8] W.H. Yang, K. Chan, and P.R. Chang, "Complex-valued neural network for direction of arrival estimation," Electron. Lett., vol.30, no.7, pp.574–575, March 1994.
[9] S.R. Buenafuente and C.M. Militello, Advances in Sound Localization, ch. 1, pp.3–6, InTechWeb, 2011.
[10] Y. Zhang and Y. Ma, "Cgha for principal component extraction in the complex domain," IEEE Trans. Neural Netw., vol.8, no.5, pp.1031–1036, Sept. 1997.
[11] A. Hirose and S. Yoshida, "Generalization characteristics of complex-valued feedforward neural networks in relation to signal coherence," IEEE Trans. Neural Netw. Learn. Syst., vol.23, no.4, pp.541–551, April 2012.
[12] F. Rosenblatt, Principles of Neurodynamics, Spartan Book, 1962.

Hirofumi Tsuzuki received a B.E. degree in 2009 from the Computer Science and Engineering Department and a M.E. degree in 2011 from the Department of Scientific and Engineering Simulation, both at the Nagoya Institute of Technology. He is currently a doctoral student at the same institute. His research interests include neural networks, signal processing and machine learning.

Mauricio Kugler received the degree in electrical engineering in 2000, and the MSc degree in biomedical engineering in 2003, both from the Federal Technological University of Parana, Brazil. In 2007, he received a PhD degree in computer science and engineering from the Nagoya Institute of Technology, Japan. Currently, he is an assistant professor at the Department of Computer Science and Engineering at this same institute. His research interests include machine learning, large scale pattern recognition methods, digital signals processing, neural networks and reconfigurable hardware. He is a member of the Institute of Electrical & Electronics Engineers (IEEE).

Susumu Kuroyanagi received a B.S. in 1991 from the Department of Electrical and Computer Engineering at the Nagoya Institute of Technology. He completed the first half of the doctoral program in 1993 and the second half in 1996, receiving the D.Eng. degree from the same institute. In 1996, he became a research associate in the Department of Electrical and Computer Engineering at the Nagoya Institute of Technology, and, in 2003, a research associate in the Graduate School of Engineering, at the Department of Computer Science and Engineering. Since 2006, he has been an associate professor in this same Graduate School. He is engaged in researches about neural networks and auditory information processing, also being a member of the Acoustic Society of Japan, the Japan Neural Network Society and Japanese Society for Medical and Biological Engineering.

Akira Iwata received a B.S. in 1973 from the Department of Electrical Engineering, Faculty of Engineering, Nagoya University. He completed the M.E. program in 1975 and became a research associate in the Department of Information, Nagoya Institute of technology. He was a visiting researcher from April 1982 to October 1983 in the research Institute of Medical information, University of Giessen Medical School, Germany. He became an associate professor in the Department of Information, Nagoya Institute of Technology in 1984, and a professor in the Department of Electrical and Computer Engineering in 1993, and vice president in 2002, and has been a professor in the Department of Computer Science and Engineering, Graduate School, since 2004. He is engaged in research on neural networks and internet security, He holds a D.Eng. degree. He received an IEICEJ paper Award in 1993 and an Information Processing Society Best Author Award in 1998. He is a member of the Information Processing Society, JSMEBE, the Japan Electrocardiography Society, the Japan Neural Network Society, and Japan Society for Medical Information Processing. He is an IEEE Senior Member.