PAPER *Special Section on Digital Signal Processing*

A Study on Precursor Signal Extraction with PCA for Predicting Significant Earthquakes

Shinji NIWA[†a)], ***Student Member***, **Hiroshi YASUKAWA**[†b)], **Ichi TAKUMI**[††c)], ***and*** **Masayasu HATA**[†††d)], ***Regular Members***

SUMMARY The tectonic activities that precede significant earthquakes release electromagnetic (EM) waves that can be used as earthquake precursors. We have been observing EM radiation in the ELF (extremely low frequency) band at about 40 observation stations in Japan for predicting significant earthquakes. The recorded signals contain, however, several noise components generated from the ionosphere, human activity, and so on. Most background noise in observed signal is attributed to lightning in the tropics. This paper proposes method based on PCA (principal component analysis) to extract signals from large data sets. The good performance of the proposed method is confirmed.

key words: *signal extraction, PCA, electromagnetic wave, ELF band, earthquake*

1. Introduction

It is well known that EM (electromagnetic) waves are radiated from the earth's surface before earthquakes [1]–[4], [9], [10]. In addition, some significant methods to predict an earthquake are also known [4]–[8], [11]–[14]. Therefore, we have been observing EM radiation in the ELF (extremely low frequency) band at over forty sites in Japan. Our goal is to be able to use EM precursor signals to predict earthquakes. The ELF band is suitable for detecting the weak precursor signals. Several investigations on the detection of precursor signals have been reported in [11], [12].

The main problem to uncertainty in ascribing the origin of a signal's fluctuation. This is because the observed signals are contaminated with several noise components such as lightning radiation, man-made noise, and so on. In this paper, we use PCA (principal component analysis) to create a method that can extract precursor signals from large data sets.

[14] reported on PCA of ULF (ultra low frequency) geomagnetic data. In [14], while, the first and second principal components are found to be associated with geomagnetic variation and man-made activity, respectively, the third principal component (residual noise like seismogenic emission if it exists) has indicated.

In our proposed system, it has a possibility that the first principal component indicates EM precursor signal. We also demonstrate the effectiveness of the method using data collected before an eruption of Mount Oyama on Miyake Island.

2. Observing Electromagnetic Radiation

This section describes the current observation system. It is important to remove several noise components when initially collecting the data. For this purpose, the observation window is restricted to the Extremely Low Frequency (ELF) band. Wide band observation might, in general, prove to be accurate. However, that many noise components lie outside the ELF band, this restriction is reasonable and provides relatively high signal-to-noise ratios.

Given that commercial power supply systems in Japan use either 50 Hz or 60 Hz, we tuned to 223 Hz (a prime number) with 1 Hz bandwidth. We used about 40 observation stations installed throughout Japan. Each observation station has three axial loop antennas with east - west, north - south and vertical orientations. The three antennas observe the variation in magnetic flux densities. The collecting circuits average the received signals over 6 second periods.

2.1 Observed Signal's Feature

The key feature of an earthquake precursor signal has been clarified in past research. Its intensity is a few to several tens of pico tesla at maximum. It increases over several days or up to 2 weeks before the earthquake. Taking this feature into account, we analyzed the observation data using DFT method [3].

Even though the observation stage eliminates many noise sources, the signal contains several other noise components. The dominant background noise is the radiation from tropical lightning, which is reflected between the ionized layer and the surface of the earth. Therefore, all data sets show a strong correla-

Manuscript received November 21, 2002.
Manuscript revised February 28, 2003.
Final manuscript received April 14, 2003.

[†]The authors are with the Faculty of Information Science and Technology, Aichi Prefectural University, Aichi-ken, 480-1198 Japan.
[††]The author is with the Faculty of Engineering, Nagoya Institute of Technology, Nagoya-shi, 466-8555 Japan.
[†††]The author is with the Faculty of Engineering, Chubu University, Kasugai-shi, 487-8501 Japan.
a) E-mail: im021018@cis.aichi-pu.ac.jp
b) E-mail: yasukawa@ist.aichi-pu.ac.jp
c) E-mail: takumi@ics.nitech.ac.jp
d) E-mail: hata@cs.chubu.ac.jp

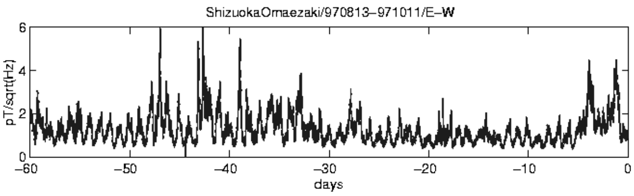

Fig. 1 Raw data captured at Omaezaki station in Shizuoka.

Table 1 Data of earthquake.

Magnitude	LNG (degree)	LAT (degree)	Depth (km)
Mj 5.0	138.2 E	34.4 N	30

tion and a daily variation. As the number of lightning events is large, the background noise is assumed to a Gaussian random noise with normal amplitude distribution statistically. There are seasonal trends; about 1 to 2 pT/$\sqrt{\text{Hz}}$ in the summer, and 0.3 to 1 pT/$\sqrt{\text{Hz}}$ in the winter.

2.2 Conventional Method

Taking account of the precursor signal's characteristic, the DFT method is often used to analyze the observed data [3]. This method extracts a low-frequency component and so eliminates most, if not all, of the noise components.

An alternative, envelope detection using band pass filtering, has been proposed [6]. In this method, the high-frequency components are extracted first, the amplitude of the component, that is, its absolute values, is calculated; it is then eliminated by a low-pass filter to leave the daily change. This method can be summarized as follows:

1. The observed signal is high-pass filtered.
2. The absolute value of the high-pass signal is calculated.
3. The absolute is out through a low-pass filter to get the envelope pattern of the high-pass signal.

Figure 1 shows a data set collected over 60 days up to October 11th, 1997, at Omaezaki station in Shizuoka. The horizontal axis represents days, and the vertical axis is the level of the observed data. The zero on the horizontal axis represents October 11th. An earthquake of magnitude Mj 5.0 occurred on that day at this area. Data of the earthquake is listed in Table 1. The low frequency component and the envelope pattern of raw data are shown in Fig. 2 and Fig. 3, respectively.

Figure 4 shows another result of the conventional method. This figure provides a one-day close-up. The horizontal axis represents hours. An earthquake of magnitude Mj4.4 occurred at Kagoshima on the 20th of April 1998. Data of the earthquake are listed in Table 2. A precursor signal was observed at Kumamoto Tomochi station on 13th of April.

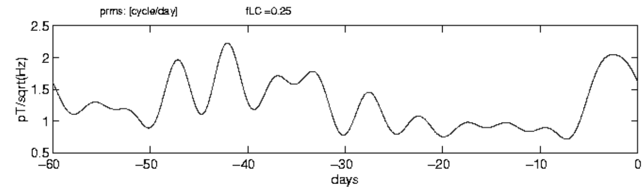

Fig. 2 Result of conventional methods. (low frequency component)

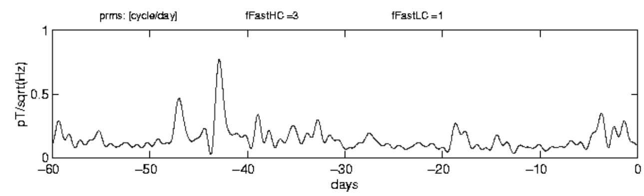

Fig. 3 Result of conventional methods. (envelope pattern)

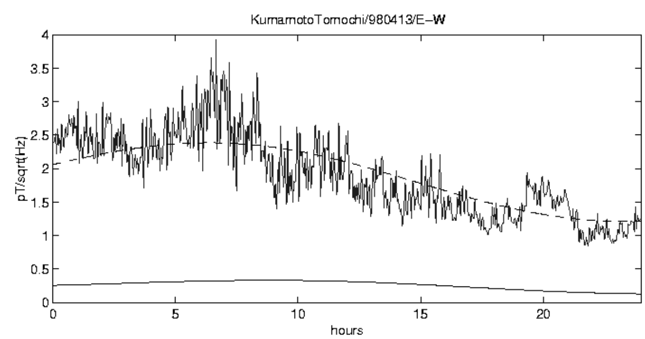

Fig. 4 An example of the signal captured at Tomochi station in Kumamoto. An earthquake occurred on October 20th, 1998, near the station.

Table 2 Data of earthquake.

Magnitude	LNG (degree)	LAT (degree)	Depth (km)
Mj 4.4	130.5 E	32.0 N	10

The conventional method cannot express the detail of the precursor signal very well. In addition, the result still contains background noise. Therefore, it can be said that the precursor signal is not really useful. This paper overcomes this weakness by using PCA to cleanly extract the true precursor signal.

3. Principal Component Analysis

PCA focuses attention on the power of the signals and the correlation between them. This method is suits this problem because the precursor signals have higher power than the steady-state signals, and the observed signals are correlated to each other. In this paper, we give due consideration to the features of the observed data and analyze data observed at the same time but at different stations to extract the precursor signals.

Let $\boldsymbol{x}$ be a data vector $\boldsymbol{x}= [\boldsymbol{x}_1, \cdots, \boldsymbol{x}_m]$ that represents the data of m–th stations $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_m$ as column vectors. Let $\overline{\boldsymbol{x}}= [\overline{\boldsymbol{x}}_1, \cdots, \overline{\boldsymbol{x}}_m]$ be the average vector of $\boldsymbol{x}$. PCA starts by calculating eigen vector $\boldsymbol{\Lambda}= \mathrm{diag}[\lambda_1, \lambda_2, \cdots, \lambda_m]$ and eigenvector $\boldsymbol{V}$ from the covariance $\boldsymbol{S}$ of the data vector $\boldsymbol{x}$ as follows,

$$\begin{aligned}\boldsymbol{S} &= \frac{1}{N}(\boldsymbol{x}-\overline{\boldsymbol{x}})'(\boldsymbol{x}-\overline{\boldsymbol{x}}) \\ &= \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}', \end{aligned} \tag{1}$$

where $(\boldsymbol{x}-\overline{\boldsymbol{x}})'$ is the transposition matrix of $\boldsymbol{x}-\overline{\boldsymbol{x}}$.

Subspaces obtained from eigenvector $\boldsymbol{V}$ represent each signal source space, and eigen values $\lambda_1, \lambda_2, \cdots, \lambda_m$ corresponding to these subspaces represent the power of each space. The principal component vector $\boldsymbol{Z}= [\boldsymbol{z}_1, \cdots, \boldsymbol{z}_m]$ is given by

$$\boldsymbol{Z} = \boldsymbol{x}\boldsymbol{V}. \tag{2}$$

where $\boldsymbol{z}_m$ is the m–th principal component.

4. Result

Figure 5 shows the precursor signal captured on April 13th by Tomochi station as well as four arbitrary one-day data sets observed at the same time at different stations. The five observation stations are listed in Table 3. The "Distance" in Table 3 is the distance between the earthquake's epicenter and each station. These signals are correlated due to the background noise.

Figure 6 and Fig. 7 show the results of PCA. The upper line and the lower line in Fig. 6 show the first and second principal component, respectively. The middle line, the bottom line and the top line in Fig. 7 show the third, fourth and fifth principal component, respectively. The precursor signal is found in only one station's data. Signals other than the precursor and the background noise are assumed to be very weak. Actually, the cumulative contribution ratio of the precursor and the background noise is large (more than 80 per cent) and constant. The cumulative contribution ratios of the result are listed in Table 4. The cumulative contribution ratio of the precursor and the background noise is about 98 per cent.

The next task is to determine what the PCA results represent; we focus on the first and second principal components in this paper since they represent the most significant information. The third to fifth principal components are very weak signals and can be ignored.

The first principal component represents a daily variation, because its waveform is similar to those of data that do not contain a precursor signal. The second component represents a reversed version of the precursor signal. Note that it is considered that the first and the second principal components contain some errors, because it is unlikely that the precursor signal and the background noise are really orthogonal in the physical

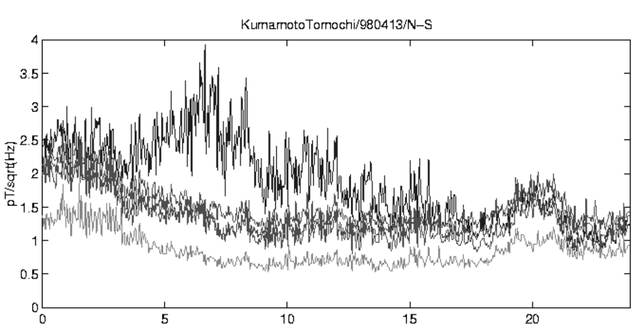

Fig. 5 An earthquake of Magnitude Mj4.4 occurred on April20th, 1998, at Kagoshima prefecture. The data captured by Tomochi station contains a precursor signal.

Table 3 Stations whose data was subjected to PCA.

Station name	Precursor	Distance (km)
ItohUsami		865.6
GifuKaidakogen		782.8
KumamotoTomochi	*	80.3
NiigataSasagami		1030.9
MiyagiWakayanagi		1220.1

*: The precursor signal is observed only at Tomochi station.

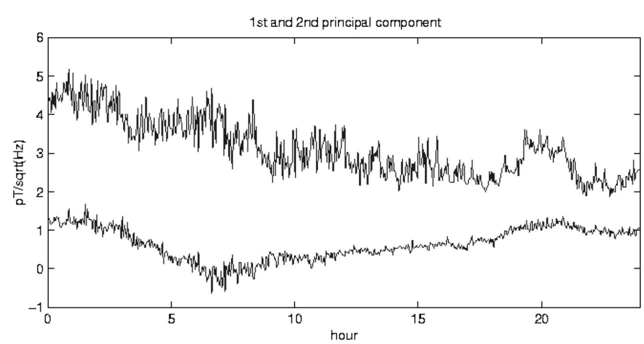

Fig. 6 The first and second principal components.

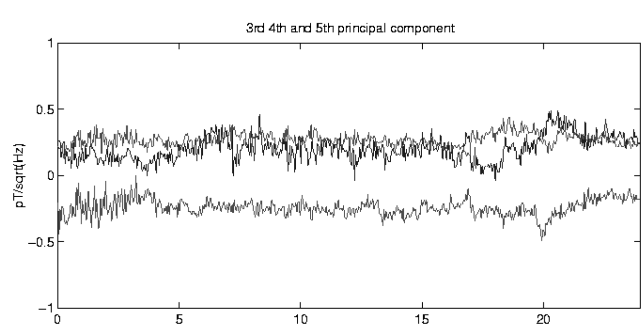

Fig. 7 The third, fourth, and fifth principal components.

Table 4 The contribution ratio.

First	Second	Third	Fourth	Fifth
0.7461	0.2336	0.0116	0.0054	0.0033

Table 5 Data of earthquakes.

Magnitude	LNG (degree)	LAT (degree)	Depth (km)
2.5 – 5.5	139.1 E	34.1 N	10 – 30

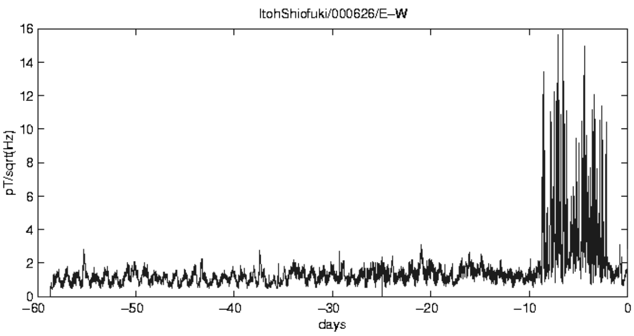

Fig. 8 The observed data contains some precursor signals. A lot of earthquakes caused by eruptions started on June 26th, 2000.

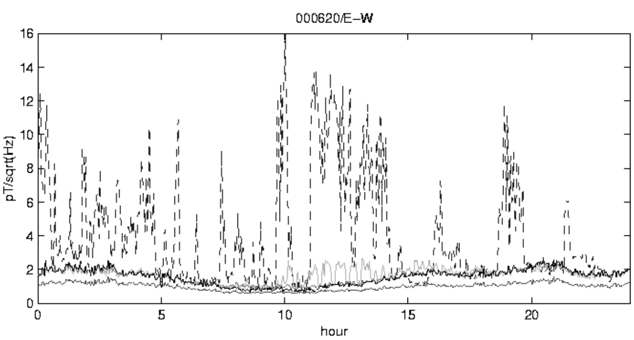

Fig. 9 Raw data observed at 5 different stations. The blue line shows Shiofuki station's data, and the other data sets show a correlation.

Table 6 Stations whose data was subjected to PCA.

Station name	Precursor	Distance (km)
ItohShiofuki	*	96.7
ItohKawana		95.3
KanagawaYugawara		119.0
ShizuokaOmaezaki		113.8
YamanashiYamanakako		149.6

*: The precursor signal was observed only at Shiofuki station.

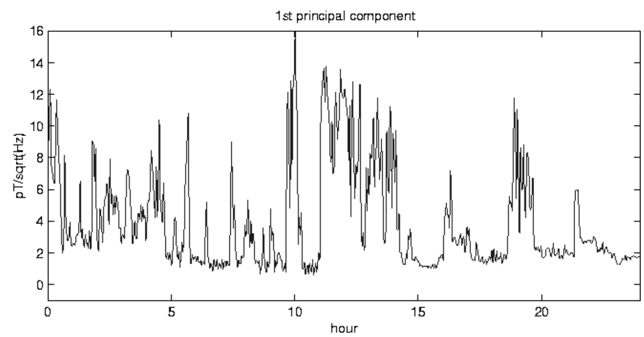

Fig. 10 The first principal component: represents the precursor signal.

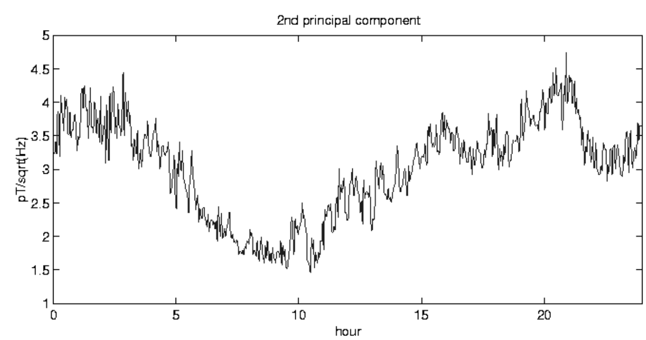

Fig. 11 The second principal component: represents the background noise.

sense.

PCA can isolate the precursor signal from the background noise. Although, the estimated signal is an antiphase version, the result is a better estimate of the precursor signal than is possible with the conventional method.

We have another example. Mt. Oyama on Miyake Island erupted many times over one month in the summer of 2000. A lot of earthquakes caused by the eruptions started on June 26th and continued for three months and more. Data of the earthquakes are listed in Table 5. Precursor signals were observed at Shiofuki station in Itoh city, Shizuoka prefecture (Fig. 8). The level of the precursor signals is about ten times larger than the level of the steady-state background signal.

In this case, the data vector $\boldsymbol{x}$ has five columns including the data containing the precursor signal. Figure 9 shows the raw data captured at the 5 stations and listed in Table 6. The data of Shiofuki station is shown as the broken line; it fluctuates drastically. On the other hand, the other signals correlate strongly to each other because they do not contain the precursor signal.

Figure 10 and Fig. 11 show the first and the second principal components, respectively. From an analysis of their shape, it seems that the first principal component represents the precursor signal and the second principal component represents the daily variation.

Note that the signal spaces yielded by the first and second principal components differ from those seen in the Kagoshima example. It is difficult to pinpoint which is the precursor signal in advance. This problem arises because the analysis focuses on signal power. The signal spaces yielded by the first and second component are dependent on the levels of the signals. Therefore, it is difficult to make a classification by using eigen values or eigenvectors.

We automatically classify the two principal components using their similarity to the daily variation de-

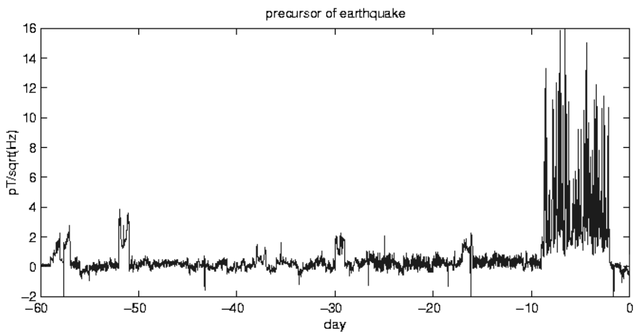

Fig. 12 Principal components classified as precursor signal.

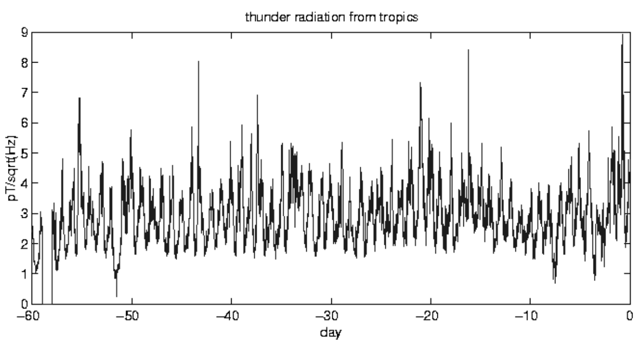

Fig. 13 Principal components classified as daily variation.

termined a priori. We adopt $\boldsymbol{u}$ as the basis for identifying the class "daily variation." $\boldsymbol{u}$ is also defined as the synchronous data of the daily variation for 15 days. This "15 day" set was selected based on our experience. The background noise shows the daily variation as mentioned before. However, this variation is not constant. In order to accommodate such changes, synchronous summation of results obtained in a suitable period should be used.

Two similarities, s_1 and s_2, between the synchronous daily variation and the first and second principal components are obtained as follows,

$$\begin{aligned} s_1 &= \cos\theta_1 \\ &= \frac{(\boldsymbol{u}, \boldsymbol{z}_1)}{\|\boldsymbol{u}\| \cdot \|\boldsymbol{z_1}\|} \end{aligned} \tag{3}$$

$$\begin{aligned} s_2 &= \cos\theta_2 \\ &= \frac{(\boldsymbol{u}, \boldsymbol{z}_2)}{\|\boldsymbol{u}\| \cdot \|\boldsymbol{z_2}\|}. \end{aligned} \tag{4}$$

The principal component that shows highest similarity is classified as the daily variation. The other is classified as the precursor signal. Principal components classified as precursor signals are shown in Fig. 12. Each principal component is only placed in an order. The precursor signal is shown at the right of the figure. On the hand, principal components classified as daily variation are shown in Fig. 13. Note that these signals are steady. Thus, the proposed method can classify the PCA output as either daily variation or an abnormal signal automatically.

5. Discussion

In this paper, we proposed a method of extracting the precursor signal from multiple data sets that is based on PCA. It yields better results than the conventional method. We presented only two cases to confirm the method's effectiveness and more case need to be tried.

The examples shown here consisted of five data sets. While it is preferable to increase the number of sets, this complicates precursor extraction because each data set contains some local noise components. A future task is to determine the optimum number of data sets. In addition, it turns out that precursors can be detected simply by means of the average of the data used in this paper. However, it is difficult to detect a precursor by using this method, since the effect of variation due to location is too large. This problem should be considered at using PCA and is foremost task in the future.

Note that the examples consisted of 1 set with precursor signal and 4 sets without. Another task is to examine signal spaces when 2 or more data sets contain a precursor signal.

6. Conclusions

We have presented a method that can extract precursor signals from EM radiation data in the ELF band. The precursor signals were accurately extracted by analyzing multiple data sets collected at the same time. In addition, we could classify the first and second components automatically by assessing their similarity to the daily variation. Further studies on the number of data sets are needed for improving the precision of this method.

Acknowledgment

This work has been supported by the Grant-in-Aid for Scientific Research (No. 11305031), the Ministry of Education, Culture, Sports, Science and Technology, Japan.

References

[1] M. Hata, I. Takumi, and S. Yabashi, "A model of earthquake seen by electromagnetic observation—Gaseous emission from the earth as main source of pre-seismic electromagnetic precursor and trigger of followed earthquake," European Geophysical Soc., NH080, 23rd General Assembly, Nice, France, 1998.

[2] B. Singh, R.P. Singh, V. Bansal, M. Kumar, and M. Hayakawa, "Anomalous subsurface VLF electric field changes associated with earthquakes and nuclear explosions observed at agra," J. Atmos. Electr., vol.19, no.2, pp.119–134, 1999.

[3] M. Hata, I. Takumi, S. Adachi, and H. Yasukawa, "An analytical method to extract precursor from noisy atmospherics," Proc. European Geophysical Society XXV, NH014, pp.25–29, Nice, France, 2000.
[4] H. Yasukawa, S. Adachi, I. Takumi, and M. Hata, "Signal sonification of seismic electromagnetic radiation in ELF band," IEEE Geoscience and Remote Sensing Sympo., Proc. IEEE Geoscience and Remote Sensing Sympo., pp.301–303, July 2000.
[5] K. Gotoh and M. Hayakawa, "A proposal on the application of blind separation for identification of seismo—ULF emissions under the noisy environment," IEICE Trans. Fundamentals (Japanese Edition), vol.J83-A, no.12, pp.1477–1485, Dec. 2000.
[6] S. Adachi, H. Yasukawa, I. Takumi, and M. Hata, "Signal processing and sonification of seismic electromagnetic radiation in the ELF band," IEICE Trans. Fundamentals, vol.E84-A, no.4, pp.1011–1016, April 2001.
[7] I. Takumi, M. Nagai, S. Kato, M. Hata, and H. Yasukawa, "Estimation of EM source location from 223 Hz EM field power data at multiple spots," Proc. Int. Geoscience and Remote Sensing Sympoium (IGARSS), 2001.
[8] H. Yasukawa, S. Adachi, M. Hata, and I. Takumi, "Signal detection and processing of seismic electromagnetic radiation in ELF band," The 8th IEEE Int. Conf. Electronics, Circuits and Systems, pp.1477–1480, 2001.
[9] M. Hata, H. Yasukawa, I. Takumi, and S. Yabashi, "ELF radiation due to diastrophism of magma dike growth," EGS XXVI General Assembly Millennium Conf. on Earth, Planetary & Solar System Science, PG9-26, 2001.
[10] S. Uyeda, M. Hayakawa, T. Nagao, O. Molchanov, K. Hattori, Y. Orihara, K. Gotoh, Y. Akinaga, and H. Tanaka, "Electric and magnetic phenomena observed before the volcano-seismic activity in 2000 in the Izu Island region, Japan," Proc. Natl. Acad. Sci. USA, vol.99, Issue 11, pp.7352–7355, May 2002.
[11] S. Niwa, H. Yasukawa, M. Hata, and I. Takumi, "Signal detection on precursor of earthquake using normal value of ELF electromagnetic wave observation," ISITA, pp.863–866, Oct. 2002.
[12] H. Yasukawa, S. Niwa, M. Hata, and I. Takumi, "On signal processing for seismic electromagnetic wave radiation application to great earthquake prediction," 2002 ICFS, pp.12-15–12-18, 2002.
[13] I. Takumi, S. Murakami, A. Shimura, M. Hata, and H. Yasukawa, "Clustering of radiation source based eigenvector," Proc. Int. Geoscience and Remote Sensing Symposium (IGARSS), INT1_D17_01, 2002.
[14] K. Gotoh, Y. Akinaga, M. Hayakawa, and K. Hattori, "Principal component analysis of ULF geomagnetic data for Izu islands earthquakes in July 2000," J. Atmos. Electr., vol.22, no.1, pp.1–12, 2002.

Shinji Niwa was born in Aichi, Japan, on February 24, 1980. He received the B.E. degree in information science and technology from Aichi Prefectural University in 2002. He is currently a candidate for the M.E. degree at Aichi Prefectural University. His research interests include digital signal processing.

Hiroshi Yasukawa received B.E., M.E. and Ph.D. degrees in electrical and electronics engineering from Shizuoka University, Hamamatsu, in 1970, 1972, and 1993, respectively. He joined the Nippon Telegraph and Telephone Public Corporation in 1972. He worked on the research and development of analog and digital communication systems in NTT Laboratories. Since April 1998, he has been a professor of Aichi Prefectural University. His research interests include digital signal processing, communication systems, and information networks. Dr. Yasukawa is a member of the IEEE, the Information Processing Society of Japan, the Acoustical Society and the European Association for Speech, Signal and Image Processing (EURASIP).

Ichi Takumi received B.E. and M.S. degrees from Nagoya Institute of Technology, Nagoya, Japan in 1982 and 1984, respectively, both in electronic engineering. After graduation he joined Oki Electric Co. He has a Doctor of Eng. degree from Nagoya Institute of Technology. Since December 1985, he has been with the Nagoya Institute of Technology, where he is now an Associate Professor in the Department of Artificial Intelligence and Computer Science. His current research interests include digital signal processing and digital communications. He is a member of the Society of Instrument and Control Engineering of Japan.

Masayasu Hata graduated in 1958 from the Department of Electronic Engineering, Faculty of Engineering, Nagoya Institute of Technology, and is affiliated with Oki Electric Co. He has a Doctor of Eng. degree from Tokyo Institute of Technology. He was engaged in R&D of digital communication system, application of electronic circuits and millimeter wave communication equipment. Upon retiring from Oki Electric Co. in 1985, he became a Professor in the Department of Artificial Intelligence and Computer Science, Nagoya Institute of Technology. He joined Aichi Prefectural University in 1998 and since 2002 he has been with Chubu University as a Professor. He is engaged in research on digital signal processing and information communication.