http://www.zju.edu.cn/jzus; E-mail: jzus@zju.edu.cn
ISSN 1009 – 3095 *Journal of Zhejiang University SCIENCE* V.4, No.5, P.595 – 601, Sep. – Oct., 2003

Extracting invariable fault features of rotating machines with multi-ICA networks*

JIAO Wei-dong(焦卫东)†, YANG Shi-xi(杨世锡), Wu Zhao-tong(吴昭同)
(*Department of Mechanical Engineering, Zhejiang University, Hangzhou 310027, China*)
†E-mail: lzn_jwd@cmee.zju.edu.cn

Received Nov.20, 2002; revision accepted Jan.16, 2003

Abstract: This paper proposes novel multi-layer neural networks based on Independent Component Analysis for feature extraction of fault modes. By the use of ICA, invariable features embedded in multi-channel vibration measurements under different operating conditions (rotating speed and/or load) can be captured together. Thus, stable MLP classifiers insensitive to the variation of operation conditions are constructed. The successful results achieved by selected experiments indicate great potential of ICA in health condition monitoring of rotating machines.

Key words: Independent Component Analysis (ICA), Mutual Information (MI), Principal Component Analysis (PCA), Multi-Layer Perceptron (MLP), Residual Total Correlation (RTC)
Document code: A **CLC number:** TN912.3

INTRODUCTION

Rotating machines such as electromotor, dynamotor, turbo compressor, etc. are important equipments in many industry fields. People had been paying considerable attention to their condition monitoring and fault diagnosis (Xu, 1998). Theoretically, any fluctuation of forces and/or movements (caused by e.g. shaft imbalance, impacts due to bearing faults etc.) can be looked on as a significant source (Lyon, 1987), in which valuable information for health condition monitoring is embedded. Health condition monitoring of rotating machines can, in essence, be looked on as a kind of special pattern recognition, which involves feature extraction of fault modes (for example imbalance, misalignment, loose foundation, etc.) expressed by multidimensional vibration measurements. However, fault-related machine vibration is usually corrupted with structural machine vibration and noise from interfering machinery. Moreover, sensors that are distributed on the machine casing will be able to measure only a mixture of the underlying vibration sources because of spatial redundancy and diversity: many fault-related peaks in the spectrum will be visible at several sensors at the same time (Alexander *et al*., 2002), which makes effective feature extraction difficult. Principal Component Analysis (PCA) for feature extraction was examined in (Chakra and Allan, 2000; Lee, 1999). PCA represents the data in orthogonal basis determined merely by the second order statistics, which is adequate for gaussian data analysis (Chitroub *et al*., 2001). However, it is well known that incipient faults in a rotating machine manifest themselves as impulses in the time signal, e.g. it had been reported that bearing failure and loose foundation gave rise to nongaussianity (increased kurtosis) of vibration measurements that contained a lot of additional information in the higher order statistics (Alexander *et al*., 1999). Thus, there is often urgent requirement for new techniques and methods for nongaussian data analysis in fault diagnosis of machines.

There was recently much interest in Independent Component Analysis (ICA) methods for sources separation and feature extraction of speech, image and biomedicine signals (Porrill, 2000; Vigario *et al*., 2000; Hansen, 2000; Chitroub, 2001; Bell *et al*., 1995; Govindan

* Project supported by the National Natural Science Foundation (No. 50205025) and the Natural Science Foundation of Zhejiang Province (No. 5001004)

et al., 1998). Especially, it has been applied to mechanical sources separation of rotating machines (Alexander *et al*., 1999; 2002; Gelle *et al*., 2000). In fact, ICA is also a powerful tool for analyzing data, especially nongaussian data. As the extension of standard PCA to higher order statistics, ICA imposes statistical independence on the extracted components and has no orthogonality constraint. Whereas PCA can only impose independence up to the second order while constraining the direction vectors to be orthogonal. Statistically, ICA is a redundancy reduction technique that ensures that the Mutual Information (MI) between the filtered output channels, a measure based on all higher order statistics of the signal, is zero (Govindan and Deng, 1998), which makes ICA more powerful for nongaussian data analysis. In this paper, we will demonstrate that ICA is potentially a useful tool in automated health condition monitoring of rotating machines. In section 2, the principle of ICA, along with an information-maximization approach for ICA (Bell and Sejnowski, 1995) will be briefly described. The whole framework of our new method for health condition monitoring of rotating machines is described in Section 3. It consists of novel multi-layer neural networks, which combine several ICA networks for feature extraction, followed by a Multi-Layer Perceptron (MLP) for the final classification. The new method based on ICA for health condition monitoring of rotating machines is examined in Section 4. It is shown that ICA can effectively capture the invariable features hidden in multi-dimensional vibration observations. Conclusion and discussion are given in the Section 5.

AN INFORMATION-MAXIMIZATION ALGORITHM FOR ICA

In ICA, the measured samples (here multi-dimension vibration measurements) are thought to be linear mixtures of some underlying sources (Bell and Sejnowski, 1995). The goal of ICA is to try to find how the measured signals $\boldsymbol{x}$ are formed from the underlying signals $\boldsymbol{s}$, assuming that the signals $\boldsymbol{s}$ are statistically as independent as possible. In practice, the determination of an ICA-basis leads to an estimation problem

$$\boldsymbol{y} = \boldsymbol{s} = \boldsymbol{W}\boldsymbol{x}. \tag{1}$$

Where the observation vector $\boldsymbol{x}$ is known, but both the base vector matrix $\boldsymbol{W}$ and the coordinates $\boldsymbol{s}$ are unknown. During the estimation procedure a large amount of observation vector $\boldsymbol{x}$ from the data class of interest is needed. As a measure of statistical independence based on all higher order statistics, the Mutual Information

$$I(\boldsymbol{s}) = \int f(\boldsymbol{s}) \log \frac{f(\boldsymbol{s})}{\prod_i f_i(\boldsymbol{s}_i)} \mathrm{d}\boldsymbol{s} \tag{2}$$

or some variant of it is selected as the criterion function of the estimation problem. Where $f(\boldsymbol{s})$ is the joint distribution whose dimensionality is usually high. The marginal distributions $f_i(\boldsymbol{s}_i)$ are the coordinate distributions related to base vectors $\boldsymbol{b}_i$ which form the base vector matrix $\boldsymbol{W}$. The quantity Eq.(2) has the minimum at $f(\boldsymbol{s}) = \prod_i f_i(\boldsymbol{s}_i)$, i.e. when the coordinates $\boldsymbol{s}_i$ are statistically independent. Solving the estimation problem Eq.(1) as presented above requires the data to be nongaussian (in the case of gaussian data PCA gives the desired answer). Indeed, the linear approach of Eq.(1) often leads to algorithms that maximize nongaussianity of the recovered components, i.e. the distribution of the projection of the data onto a vector in the ICA-basis should be as far from gaussian as possible. If it cannot be assumed that independent sources underlie a dataset, ICA may therefore be regarded as a form of exploratory projection pursuit, by which 'interesting' structure in the observed space can be sought via nonlinear functions of linear projections, meanwhile, dimensionality reduction is implemented.

The basic problem tackled here is how to minimize the Mutual Information that the output $\boldsymbol{y}$ of a neural network processor contains about its input $\boldsymbol{x}$. This is defined as

$$I(\boldsymbol{y}, \boldsymbol{x}) = H(\boldsymbol{y}) - H(\boldsymbol{y} \mid \boldsymbol{x}). \tag{3}$$

Where $H(\boldsymbol{y})$ is the entropy of the output, while $H(\boldsymbol{y} \mid \boldsymbol{x})$ is whatever entropy the output has that did not come from the input. Bell and Sejnowski (1995) proposed an ICA algorithm in which they maximized the joint entropy, $H[g(\boldsymbol{W}\boldsymbol{x})]$, of the elements of the linear transform squashed by a sigmoid function, $g(\cdot)$. This sigmoid function is the cumulative density function (c.d.f.) of the signal we are trying to extract. Even if the particular c.d.f. is unknown, good results have been obtained for high kurtosis distributions by using the logistic function

$$g(\boldsymbol{u})=(1+e^{-\boldsymbol{u}})^{-1}, \quad \boldsymbol{u}=\boldsymbol{W}\boldsymbol{x}+\boldsymbol{w}_0. \tag{4}$$

Where $\boldsymbol{W}$ is the weight matrix for the ICA network, and $\boldsymbol{w}_0$ is a bias vector. For maximizing the entropy, weights are updated incrementally according to the gradient of the entropy. Thus, the resulting learning rules are

$$\Delta \boldsymbol{W} \propto [\boldsymbol{W}^{T}]^{-1}+(1-2\boldsymbol{y})\boldsymbol{x}^{T}, \tag{5}$$

$$\Delta \boldsymbol{w}_0 \propto 1-2\boldsymbol{y}, \tag{6}$$

For an individual weight, $\boldsymbol{w}_{ij}$, this rule amounts to

$$\Delta \boldsymbol{w}_{ij} \propto \frac{\mathrm{cof}\,\boldsymbol{w}_{ij}}{\det W}+\boldsymbol{x}_j(1-2\boldsymbol{y}_i). \tag{7}$$

In machine condition monitoring it is not possible to collect an exhaustive set of measurements from all possible failure scenarios. Moreover, since a machine may be used under very different operating conditions (running speed, load, type of lubricants) and environmental conditions (indoor or outdoor, placed near interfering machinery), each fault mode actually comprises a set of patterns (Alexander *et al.*, 2002). Consequently, a set of feature vectors have to be used for the description of a fault pattern, which leads to high computational complexity and difficulty for final classification. It will be seen in the next section that such difficulty can be overcome to some extent by the use of multi-ICA networks for the feature extraction.

FEATURE EXTRACTION AND CLASSIFICATION

The block diagram of our new classification system is given in Fig. 1. The logic behind our method is very similar with the one by Govindan *et al.* (1998), which applied ICA to Electrogram classification during Atrial Fibrillation. When multidimensional vibration measurements belonging to one fault class is passed through an ICA network which has been trained on measurements from that particular fault class, the residual correlation between the output channels would be less than when it is passed through a network trained on a different class. Here, three faults i.e. imbalance, impact and loose foundation are induced to a rotor kit* separately. This experiment setup will be introduced in next section.

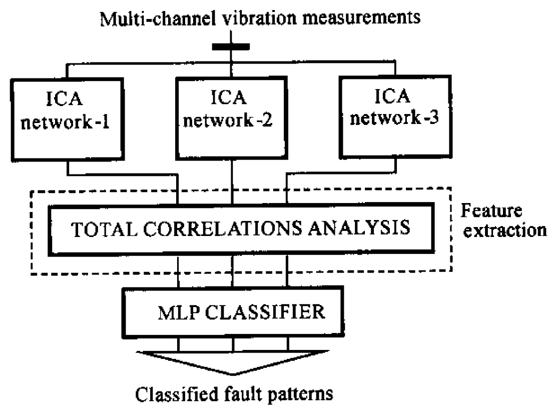

Fig. 1 Block diagram of the compound ICA-MLP networks inspired by Govindan *et al.* (1998)

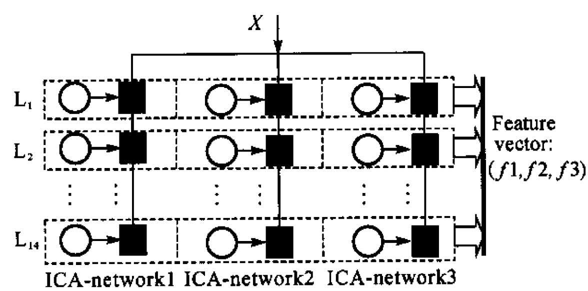

Fig. 2 Feature extraction with multiple layer ICA networks

We select data segments from each fault class under different running speed and load to build up the training set which was used to batch train three ICA networks, each one being trained on one class of data. The 'total correlations' at the output of the three ICA networks, when a data segment was presented at the input, formed the feature vector for that data segment, which was used to train a Multiple Layer Perceptron (MLP), which performed the final classification. The 'total correlations' at the output of one ICA network can be defined as follows: (Govindan and Deng, 1998)

$$totalcorrelation = \sqrt{\sum_{i,j=1}^{N} corcoefficient_{ij}^{2}}. \tag{8}$$

Where $corcoefficient_{ij}$ is the correlation coefficient

* Produced by BENTLY Nevada Co., USA

at zero lag between the i output channel and the j output channel. To test the algorithm, each data segment from the test set was presented at the input of the three ICA networks. The output of the MLP indicates which ICA network gave the minimum residual correlation between channels, which in turn indicates the fault class to which the vibration measurements belongs. The total procedure of ICA feature extraction can be described in Fig.2, where ○ and ■ stand for the training set by which the 'independence' of data is found and the trained ICA network, respectively. X is the data segment to be tested, and L1, L2...L14 stand for different shaft rotating speeds from 500r/min to 7000r/min. At last, a feature vector ($f1$, $f2$, $f3$) can be extracted for every data segment obtained at certain rotating speed when it is passed through the corresponding trained ICA network.

EXPERIMENT

The setup used for fault simulation and signal acquisition is described in the Fig.3.

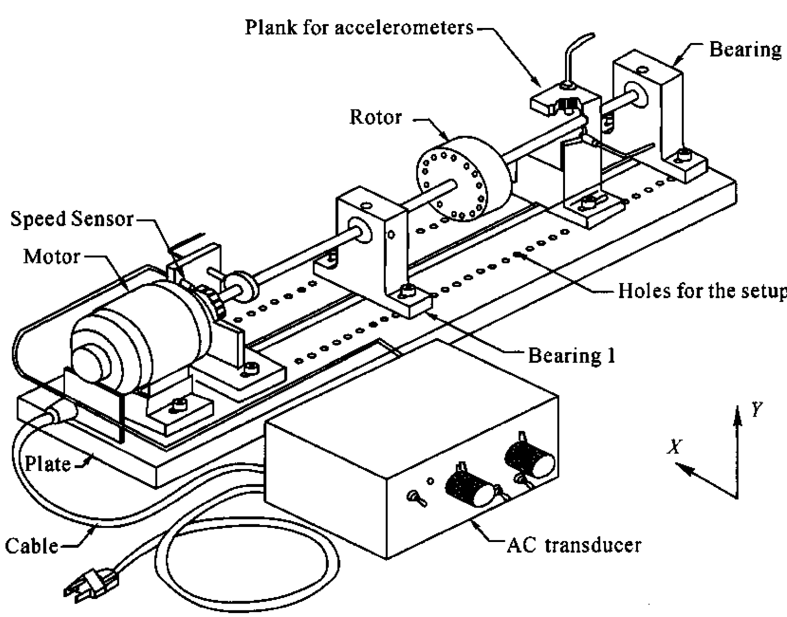

Fig.3 The Bently rotor kit

By the use of the AC transducer in Fig. 3, the motor can be made to run at several shaft-rotating speeds (from 200r/min to 8000r/min). A number of faults, for example Imbalance, Impact and Loose Foundation, etc. were induced in this rotor kit, and vibration was measured with two accelerometers fixed on the plank in two directions X and Y. Both normal and faulty behavior was measured at several speeds. However, only the fault observations from 500 r/min to 7000 r/min shaft-rotating speeds were used in this paper. The strategy of sampling the vibration signals is equiv-cyclic by means of the phase signal collected by the speed sensor.

In order to evaluate clearly and synthetically the capacity of multi-ICA networks for feature extraction, vibration signals were recorded in two channels from every fault pattern (i.e. Imbalance, Impact and Loose Foundation) under different rotating speeds ranging from 500 r/min to 7000 r/min. The vibration data were filtered using band-pass amplifiers and digitized at 64 spectral bins. Every fault observation at every rotating speed consists of one hundred data segments, each segment 512 samples long. We analyzed these data using the information-maximization approach for ICA by Bell and Sejnowski (1995) and Netlab toolbox, which is available free at web-address: http://www.ncrg.aston.ac.uk/netlab. The first fifty data segments from each fault class were selected to build up the training set used to batch train three ICA networks, each one being trained on one class of data. By this way, the results of training three ICA-networks can be obtained as shown in Fig.4.

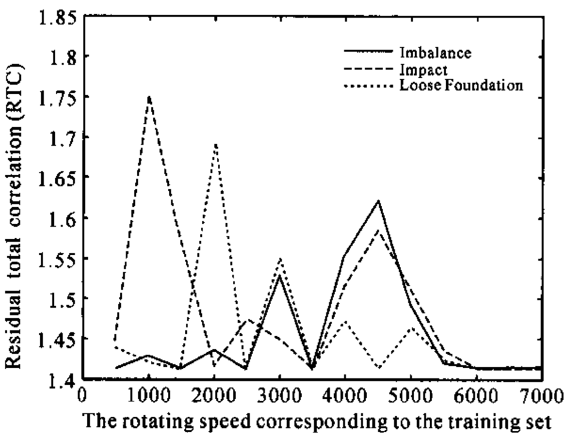

Fig.4 The RTC of three ICA-networks trained by the training set from 500r/min – 7000r/min

In Fig.4, the solid, dashed and dotted line, respectively, stands for the variation of Residual Total Correlation (RTC) of the trained ICA network1 (trained by Imbalance samples), ICA network 2 (trained by Impact samples) and ICA network 3 (trained by Loose Foundation samples). It can be seen that every curve fluctuates violently, which implies that the inherent data structure captured by ICA varies with respect to the different shaft-rotating speed under which different training segments were collected by accelerometers. Subsequently, we presented the training set (i.e. the first fifty data segments in every observation) and the test set (i.e. the remaining fifty data segments in every observation) as the input of the three trained ICA-networks one by one, so that a three dimensional feature vector ($f1$, $f2$, $f3$) was outputted (See Fig. 2). According to the logic interpreted in Section 3, one can directly classify a data segment into one of the three given faults from the quantity of every variable in the feature vector extracted by the three ICA networks. As a result, we can obtain the classification error of the training set and test set in Fig.5. It can be noticed that all data segments of both the training and the test set are classified to some extent, even though there exists such violent variation in the data structure (See Fig. 4). Thus, we can also say that the invariable feature embedded in every fault mode is extracted to some degree by the multi-ICA-networks.

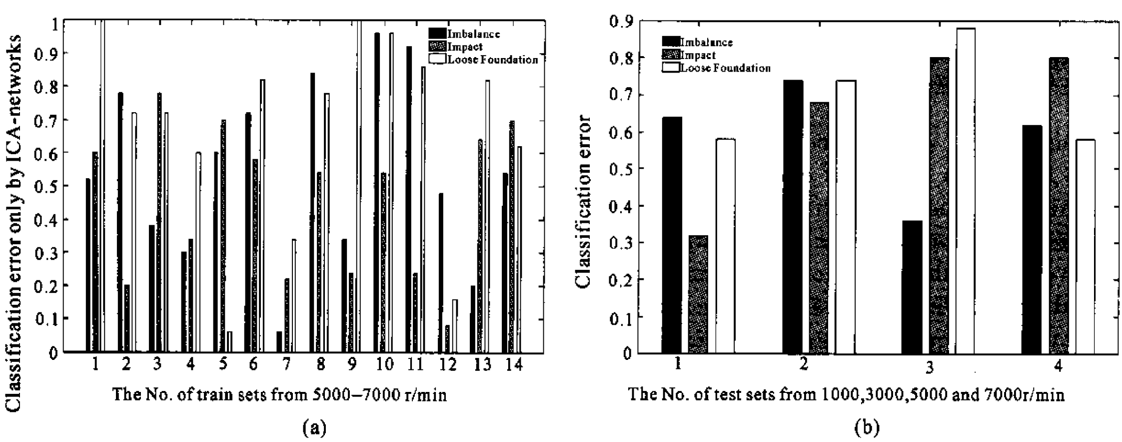

Fig.5 Classification error for the train and test segments only by the trained ICA-networks
(a) with the training segments from 500r/min – 700r/min; (b) with the test segments under 1000r/min, 3000r/min, 5000r/min and 7000r/min

Ideally, the output of the ICA networks should have three patterns corresponding to the three fault classes. But in practice, some vibration measurements gave minimum total correlation at the output of ICA networks not trained on its class. Also, the overlap between different fault classes is inevitable. Thus, minimum residual correlation cannot always be obtained by some test samples even if the ICA network had been trained by other samples that belong to the same class as the test set. So further classification by a MLP became necessary. We trained a two-layer perceptron using the feature vectors extracted by the fifty data segments in the training set. Here, the scaled conjugate gradient algorithm and a log-sigmoid transfer function were used for both layers with the number of neurons in the hidden layer being ten. Experimentally, the convergence of the MLP was obtained after less than 50 iteration steps. To test our method, each data segment from the test set was presented at the input of the three ICA networks. Also, we compared this method with one MLP classifier based on FFT features. It is well known that faults in rotating machines will be visible in the acceleration spectrum as increased harmonics of running speed or presence of sidebands around characteristic (structure-related) frequencies. We focused on relatively low feature dimensionality (64 spectral bins) during FFT feature extraction. The whole performance of our method can be summed as Table 1. Clearly, notable improvement was obtained by the use of the ICA-MLP classifier. Also, the feature dimensionality used in our ICA-MLP classifier is three, just equal to the amount of the faults (i.e. Imbalance, Impact and Loose Foundation) to be recognized. By means of the multi-ICA networks, the hidden fault information in vibration observations under different shaft-rotating speed was effectively extracted, and expressed with very low feature dimensionality. Thus, a robust MLP classifier was constructed using the training set with relatively small size, which is, clearly, important for health condition monitoring of rotating machines.

Table 1 The whole performance of ICA-MLP compared with FFT-MLP

	Imbalance	Impact	Loose Foundation
ICA-MLP trained by samples under 1000r/min	90%	86%	92%
ICA-MLP trained by samples under 3000r/min	84%	78%	96%
ICA-MLP trained by samples under 5000r/min	78%	82%	94%
ICA-MLP trained by samples under 7000r/min	84%	88%	92%
FFT-MLP using samples under 5000r/min	76%	80%	88%

DISCUSSION AND CONCLUSION

In this paper, we used compound ICA-MLP networks for health condition monitoring of rotating machines. Usually, high spectral resolution may be required for adequate fault identification due to overlap in the series of harmonic components and noise. This may lead to difficulties because of the curse of dimensionality: one needs large sample sizes in high-dimensional spaces in order to avoid overfitting of the training set. Such problem can be remedied by the use of multi-ICA networks for feature extraction. First, the multi-ICA networks can directly classify three fault modes to some extent, which implies the ability of ICA to capture the inherent features hidden in multi-dimensional vibration data. Second, satisfactory fault identification can be implemented using the MLP classifier trained by the 3-dimensional feature vectors from three ICA-networks during the faults classification. It is remarkable that the classification accuracy for three fault modes by using the trained ICA-MLP classifier is insensitive to certain a shaft-rotating speed, that is, the ICA-MLP trained by the samples under one running speed is able to classify with considerable accuracy the test set under another running speed, which shows that invariable features in a certain fault mode under different running speed can be extracted by multi-ICA networks together.

In a rotor kit, different fault modes often manifest themselves as strong nonlinear vibration. Theoretically, ICA implements feature extraction by finding the direction of the meaningful structure in data, which is usually constrained to 'independence' and 'nongaussianity' of data. And, conventional ICA assumes that observations are linear mixtures of some underlying sources, and that there is no additive noise just in (1). Indeed, some of the constraints are violated in practical applications. As a matter of fact, many faults of rotating machines (for example imbalance, oil whirl and shaft crack) often manifest themselves as strong nonlinear behaviors (Alexander *et al*., 2002). In order to relax the assumptions, new ICA mixtures need to be used. Fortunately, nonlinear ICA methods that involve nonlinear unmixing or inference solutions have been developed (Lee, 2000; Luis, 1999). It is hoped that more novel and significant features buried in multi-dimensional vibration measurements from a rotating machine can be extracted by means of appropriate nonlinear ICA algorithms, all of which are to be researched in future.

References

Alexander, Y., David, M. J. and Robert, P. W., 1999. Robust Machine Fault Detection with Independent Component Analysis and Support Vector Data Description. Pattern Recognition Group, Dept. of Applied Physics, Delft University of Technology. Delft, The Netherlands.

Alexander, Y., Amir, L. and Robert, P. W., 2002. Blind Separation of Rotating Machine Sources: Bilinear Forms and Convolutive Mixtures. Neurocomputing-Special Issue on ICA/BSS, p.123 – 134.

Bell, A. and Sejnowski, T., 1995. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7:1129 – 1159.

Chakra, Ch. and Allan, J., 2000. Sparse PCA Extracting Multi-Scale Structure From Data. Department of Computer Science, University of Toronto, 6 King's College Road Toronto, ON M5S 3H5, Canada.

Chitroub, S., Houacine, A. and Sansal, B., 2001. Compound PCA-ICA Neural Network Model for Enhancement and Feature Extraction of Multi-Frequency Polarimetric SAR Imagery. Signal Processing Laboratory, Electronics Institute, University of Sciences and Technology of Houari Boumedienne.

Gelle, G., Colas, M. and Delaunay, G., 2000. Blind sources separation applied to rotating machines monitoring by acoustical and vibrations analysis. *Mechanical Systems and Signal Processing*, **14**(3):427 – 442.

Govindan, A. and Deng, G., 1998. Independent Component Analysis applied to Electrogram Classification during Atrial Fibrillation. School of Electronic Engineering, La University, Melbourne Vic-3083, Australia.

Hansen, L. K., 2000. Blind Separation of Noisy Image Mixtures, Based on a Workshop Held after the 1999 International Conference on Artificial Neural Networks (ICANN), Kluwer Academic Publishers.

Lee, S.J., Jung, S.B., Kwon, J.W. and Hong, S.H., 1999. Face Detection and Recognition using PCA. IEEE TENCON.

Lee, T.W., 2000. Nonlinear Approaches to Independent Component Analysis. The Salk institute, Institute for Neural Computation, University of California, San Diego, USA.

Luis, B. Almeida., 1999. ICA of Linear and Nonlinear Mixtures based on Mutual Information. IST and INESC-ID, Lisbon, Portugal.

Lyon, R.H., 1987. Machinery Noise and Diagnostics. Butterworths, Boston, p.342 – 345.

Porrill, J., Stone, J.V., Berwick, J., Mayhew, J. and Coffey, P., 2000. Analysis of Optical Imaging Data Using Weak Models and ICA. Based on a Workshop Held after the 1999 International Conference on Artificial Neural Networks (ICANN), Kluwer Academic Publishers.

Vigario, R., Sarela, J. and Oja, E., 2000. Searching for Independence in Electromagnetic Brain Waves, Based on a Workshop Held after the 1999 International Conference on Artificial Neural Networks (ICANN), Kluwer Academic Publishers.

Xu, M. , 1998. Manual for Machinery Fault Diagnosis. Xi'an Jiaotong University Press, Xi'an City (in Chinese).