Proceedings of the IASTED International Conference
CIRCUITS, SIGNALS, AND SYSTEMS
May 19-21, 2003, Cancun, Mexico

INDEPENDENT COMPONENT ANALYSIS OF VIBRATIONS FOR FAULT DIAGNOSIS OF AN INDUCTION MOTOR

Sanna Pöyhönen[1)], Pedro Jover[2)], Heikki Hyötyniemi[1)]

[1] Control Engineering Laboratory,
Helsinki University of Technology
P.O.Box 5400, 02015 HUT
Finland

[2] Laboratory of Electromechanics
Helsinki University of Technology
P.O. Box 3000, 02015 HUT
Finland

ABSTRACT

ICA is applied to multi-channel vibration measurements of a 35 kW cage induction motor to fuse the information of several channels, and provide a robust and reliable fault detection routine. Independent components are found from the measurement data set with FastICA algorithm, and their PSD estimates are calculated with Welch's method. A SVM based classification routine is applied to the PSD estimates to perform the fault identification. Similar classification is applied directly to vibration measurements. Based on the results with real measurement data it is shown that data fusion with ICA enhances the fault diagnostics routine.

KEY WORDS
Electrical Machines, Vibration Monitoring, Fault Classification, Data Fusion, Independent Component Analysis, Support Vector Machine

1. Introduction

Vibration analysis has been used in rotating machines fault diagnosis for decades [1-3]. In [3], it is claimed that vibration monitoring is the most reliable method of assessing the overall health of a rotor system. Machines have complex mechanical structures that oscillate and coupled parts of machines transmit these oscillations. This results in a machine related frequency spectrum that characterizes healthy machine behaviour. When a mechanical part of the motor either wears or breaks up, a frequency component in the spectrum will change. In fact, each fault in a rotating machine produces vibrations with distinctive characteristics that can be measured and compared with reference ones in order to perform the fault detection and diagnosis.

Vibration monitoring system requires storing of a large amount of data. Vibration is often measured with multiple sensors mounted on different parts of the machine. For each machine there are typically several vibration signals being analysed in addition to some static parameters like load. The examination of data can be tedious and sensitive to errors. Also, fault related machine vibration is usually corrupted with structural machine vibration and noise from interfering machinery. Further, depending on the sensor position, large deviations on noise may occur in measurements.

Due to these problems intelligent compression of the multichannel measurement data may aid in data management for fault diagnostics purpose. Independent component analysis (ICA) can be used to find a structure in large amount of multivariate data. ICA may be used to compress measurements from several channels into a smaller amount of channel combinations - statistically independent components of the measurements - that could clearly indicate faults in the machine.

In this article, independent component analysis (ICA) is studied to provide a robust and reliable fault diagnostics routine for a cage induction motor. Ypma & al. have successfully utilised similar approach with application to fault diagnostics of a submersible pump in [4] and [5]. In our study, resulting independent components are further processed with Welch's power spectrum density (PSD) estimation and support vector machine (SVM) [6] based classification to obtain the knowledge of the motor condition. In literature, ICA has also been applied to separate machine vibrations from interfering vibration sources e.g. in [7]-[9].

The content of the paper is following. In Section 2, the basis of ICA is explained. In Section 3, SVM based classification is explained. In Section 4, the vibration measurement system is explained as well as results from fault classification based on vibration monitoring with ICA. In Section 5, some conclusions are made.

2. Independent Component Analysis [10, 11]

Independent component analysis is a method for finding underlying structure in multivariate data. What makes the difference between ICA and other multivariate analysis methods is that it looks for components that are both statistically independent, and nongaussian. For example in PCA, the redundancy is measured by correlations between data elements, while in ICA the idea of independence is used. Statistical independence means that the value of any one of the components gives no information of the other components. Principal components of the data are statistically not correlated and for gaussian data uncorrelated components are also independent. However, real data sets often do not follow a gaussian distribution. For example, many data sets have supergaussian distribution that means that the random variables take relatively more often values close to zero or very large.

Consider a situation where there are a number of signals emitted some physical objects or sources (e.g. a rotating machine vibration and interfering vibration sources). Assume further that there are several sensors or receivers (e.g. vibration sensors). The sensors are in different positions so that each records a mixture of the original source signals with slightly different weights. Let us denote ith recorded mixture with x_i and jth original source with s_j. The phenomenon can be described with an equation $\mathbf{x} = \mathbf{A}\mathbf{s}$, where elements of **x** are x_i and elements of **s** are s_j.

The elements a_{ij} of the matrix **A** are constant coefficients that give the mixing weights that are assumed to be unknown. **A** is called a mixing matrix. A blind source separation problem is to separate original sources from observed mixtures of the sources, while blind means that we know very little about the original sources and about the mixture. It can be safely assumed that the mixing coefficients are different enough to make matrix **A** invertible. Thus there exists a matrix $\mathbf{W} = \mathbf{A}^{-1}$ that reveals the original sources $\mathbf{s} = \mathbf{Wx}$. After this the problem is, how to estimate the coefficient matrix **W**. A simple solution to the problem can be found by considering the statistical independence of different linear combinations of **x**.

In [11], problem of finding independent components (IC) is formulated with the concept of mutual information. First differential entropy H of a random vector $\mathbf{y} = (y_1, y_2,...,y_n)$ with density $f(\cdot)$ is defined as follows:

$$H(\mathbf{y}) = -\int f(\mathbf{y}) \log f(\mathbf{y}) \mathbf{dy} \ . \tag{1}$$

A gaussian variable has the largest entropy among all variables of equal variance. Differential entropy can be normalised to get the definition of negentropy:

$$J(\mathbf{y}) = H(\mathbf{y}_{gauss}) - H(\mathbf{y}) \ , \tag{2}$$

where $\mathbf{y}_{gauss}$ is a gaussian random vector of the same covariance matrix as **y**. Negentropy is zero for gaussian variable and always nonnegative. Mutual information I of random variables y_i, $i = 1,\dots,n$ can be formulated using negentropy J, and constraining the variables to be uncorrelated:

$$I(y_1, y_2, ..., y_n) = J(\mathbf{y}) - \sum_i J(y_i) \ . \tag{3}$$

Since mutual information is the information theoretic measure of the independence of random variables, it is natural to use it as a criterion for finding the ICA transform. Thus the matrix **W** is determined so that the mutual information of the linear combinations of mixed variables **x** is minimized. Because negentropy is invariant for invertible linear transforms, **W** that minimizes the mutual information is roughly equivalent to finding directions in which the negentropy is maximized.

To use the formulation above a simple estimate for the negentropy or differential entropy need to be chosen. There are different options for this to emphasize e.g. robustness or fast convergence.

The FastICA algorithm [11] is a computationally highly efficient method for performing the estimation of independent components of time series. It uses a fixed-point iteration scheme that has been found to be 10-100 times faster than conventional gradient descent methods for ICA. In this research, a FastICA MATLAB-package developed at Laboratory of Computer and Information Science in the Helsinki University of Technology was applied**.**

3. Classification with Support Vector Machines

SVM based classification [6] is a relatively new machine learning method based on statistical learning theory presented by Vapnik. In SVM, an optimal hyperplane is determined to maximize the generalization ability of the classifier by mapping the original input space into a high dimensional dot product space called feature space. The mapping is based on a so-called kernel function. The optimal hyperplane is found in the feature space with a learning algorithm from optimization theory that implements a learning bias derived from statistical learning theory.

SVM based classifier is claimed to have better generalisation properties than for example neural network based classifiers. In addition to this, SVM based classification is interesting, because its computational efficiency does not depend on the number of features of classified entities. This property is very useful in fault diagnostics, because the number of features to be chosen to be the base of fault classification is thus not limited.

Let n-dimensional input $\mathbf{x}_i$ ($i = 1 ,...,M$) belong to Class I or Class II and associated labels be $y_i = 1$ for Class I and $y_i = -1$ for Class II. For linearly separable data, we can determine a hyperplane $f(\mathbf{x})$ that separates the data. For a separating hyperplane $f(\mathbf{x}) \geq 1$, if the input $\mathbf{x}$ belongs to positive class, and $f(\mathbf{x}) \leq -1$, if $\mathbf{x}$ belongs to the negative class. This results (5):

$$f(\mathbf{x}) = \mathbf{w} \bullet \mathbf{x} + b = \sum_{j=1}^{n} w_j x_j + b \qquad (4)$$

$$y_i f(\mathbf{x}_i) = y_i(\mathbf{w} \bullet \mathbf{x}_i + b) \geq 1, \quad \text{for} \quad i = 1, \ldots, M \, , \qquad (5)$$

where $\mathbf{w}$ is an n-dimensional vector and b is a scalar. The weighting vector $\mathbf{w}$ defines the direction of the separating hyperplane $f(\mathbf{x})$ and with $\mathbf{w}$ and b (bias) it is possible to define the hyperplane's distance from the origin.

The separating hyperplane that has the maximum distance between the hyperplane and the nearest data, i.e. the maximum margin, is called the optimal separating hyperplane. An example of optimal separating hyperplane of two datasets is presented in Fig. 1. The optimal hyperplane is perpendicular to the shortest line between border lines of two sets, and the plane and the shortest line intersect each other in the halfway of the line. The geometrical margin γ is half of the sum of the distances between arbitrary separating hyperplane and the nearest negative and positive datum ($\mathbf{x}^-$ and $\mathbf{x}^+$):

$$\begin{aligned} \gamma &= \frac{1}{2}\left(\left(\frac{\mathbf{w}}{\|\mathbf{w}\|_2} \cdot \mathbf{x}^+\right) - \left(\frac{\mathbf{w}}{\|\mathbf{w}\|_2} \cdot \mathbf{x}^-\right)\right) \\ &= \frac{1}{2\|\mathbf{w}\|_2}\left((\mathbf{w} \cdot \mathbf{x}^+) - (\mathbf{w} \cdot \mathbf{x}^-)\right) \quad . \end{aligned} \qquad (6)$$

Without loss of generality the optimal separating hyperplane can be searched among canonical hyperplanes that fulfil $\mathbf{w} \cdot \mathbf{x}^+ = 1 - b$ and $\mathbf{w} \cdot \mathbf{x}^- = -1 - b$:

$$\gamma = \frac{1}{2}\left(\left(\frac{\mathbf{w}}{\|\mathbf{w}\|_2} \cdot \mathbf{x}^+\right) - \left(\frac{\mathbf{w}}{\|\mathbf{w}\|_2} \cdot \mathbf{x}^-\right)\right) = \frac{1}{\|\mathbf{w}\|_2} \quad . \qquad (7)$$

The optimal hyperplane maximizes the geometrical margin. Thus the optimal hyperplane can be found by solving the following convex quadratic optimisation problem:

$$\begin{aligned} &\text{minimize } \frac{1}{2}\|\mathbf{w}\|^2 \\ &\text{subject to } \mathrm{y_i}(\mathbf{w} \bullet \mathbf{x}_i + b) \geq 1 \quad . \end{aligned} \qquad (8)$$

If the number of attributes of data examples is large, we can considerably simplify calculations by converting the problem with Kuhn-Tucker conditions into the equivalent Lagrange dual problem, which will be:

$$\begin{aligned} &\text{maximize } W(\boldsymbol{\alpha}) = \sum_{i=1}^{M} \alpha_i - \frac{1}{2}\sum_{i,k=0}^{M} \alpha_i \alpha_k y_i y_k \mathbf{x}_i \bullet \mathbf{x}_k \\ &\text{subject to } \sum_{i=1}^{M} y_i \alpha_i = 0, \quad \alpha_i \geq 0, \quad i = 1, \ldots, M \quad . \end{aligned} \qquad (9)$$

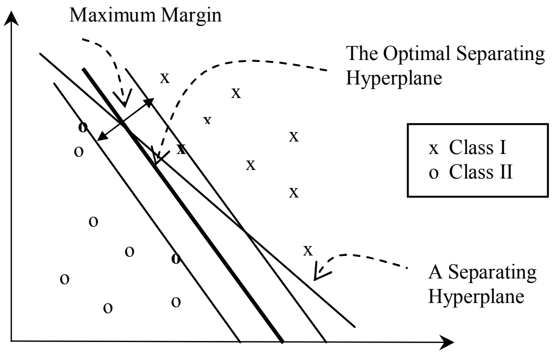

Figure 1. Optimal Hyperplane

The number of variables of the dual problem is the number of training data.

Let us assume that optimal solution for the dual problem is $\boldsymbol{\alpha}^*$ and b^*. According to the Karush-Kuhn-Tucker theorem, the equality condition in (5) holds for the training input-output pair ($\mathbf{x}_i$,y_i) only if the associated α_i^* is not 0. In this case the training example $\mathbf{x}_i$ is a support vector. Solving (9) for $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_M)$, we can obtain the support vectors for classes I and II. Then the optimal separating hyperplane is placed at the equal distances from the support vectors for classes I and II, and b^* is given by:

$$b^* = -\frac{1}{2}\sum_{k=1}^{M} y_k \alpha_k^* (\mathbf{s}_1 \bullet \mathbf{x}_k + \mathbf{s}_2 \bullet \mathbf{x}_k) \, , \qquad (10)$$

where $\mathbf{s}_1$ and $\mathbf{s}_2$ are arbitrary support vectors for Class I and Class II, respectively. In Fig.1, support vectors are bolded. Notice, that support vectors are such training samples that are on the margin of two datasets. The optimal separating hyperplane would be the same, if only support vectors had been used as training data.

So far it is assumed that the training data is linearly separable. In the case where the training data cannot be linearly separated, non-negative slack variables ξ_i are introduced to (5). This corresponds to adding the upper bound C to $\boldsymbol{\alpha}$. In both cases, the decision functions are the same and are given by:

$$f(\mathbf{x}) = \sum_{i=1}^{M} \alpha_i^* y_i \mathbf{x}_i \bullet \mathbf{x} + b^* \quad . \qquad (11)$$

Then unknown data example $\mathbf{x}$ is classified as follows:

$$\mathbf{x} \in \begin{cases} \text{Class I,} & \text{if } f(\mathbf{x}) > 0 \\ \text{Class II,} & \text{otherwise} \quad . \end{cases} \qquad (12)$$

SVM is a non-linear kernel-based classifier, which maps the data to be classified, X, onto a space, where the data can be linearly classified. Using the non-linear vector function $\mathbf{\Phi(x)} = (\Phi_1(\mathbf{x}),\ldots,\Phi_l(\mathbf{x}))$ that maps the n-dimensional input vector $\mathbf{x}$ into the l-dimensional feature space, the linear decision function in dual form is given by:

$$f(\mathbf{x}) = \sum_{i=1}^{M} \alpha_i y_i \mathbf{\Phi}(\mathbf{x}_i) \bullet \mathbf{\Phi}(\mathbf{x}) . \qquad (13)$$

Notice that in (13) as well in the optimisation problem (9), the data occur only in inner products. In SVM, the actual mapping function, **Φ**, is not necessary to be known, but the classes optimally separating hyperplane is possible to calculate with inner products of the original data samples. If it is possible to find this kind of procedure to calculate inner products of feature space in original data space, it is called a kernel, $K(\mathbf{x},\mathbf{z}) = \Phi(\mathbf{x}) \bullet \Phi(\mathbf{z})$. Then the learning in the feature space does not require evaluating **Φ** or even knowing it, because all the original samples are handled only with Gram matrices $G = ((\mathbf{x}_i \bullet \mathbf{x}_j))_{i,j=1}^{M}$. Using a Kernel function, the decision function will be:

$$f(\mathbf{x}) = sign\left(\sum_{\text{support vectors}} \alpha_i^* y_i K(\mathbf{x}_i, \mathbf{x}) \right) , \qquad (14)$$

and the unknown data example is classified as before.

However, all kernels do not correspond to inner products in some feature space. With a so called Mercer's theorem it is possible to find out, whether a kernel K depicts an inner product in that space where **Φ** is mapped [6].

Least Square Support Vector Machines (LS-SVM) are reformulations to the standard SVM's [12]. The cost function is a regularized least squares function with equality constraints leading to linear Karush-Kuhn-Tucker systems. The solution can be found efficiently by iterative methods like conjugate gradient algorithm. LS-SVM's do not lead sparse solutions like SVM's but they are computationally very fast. In this research, LS-SVM MATLAB toolbox presented in [12] is used.

4. Results

4.1 Vibration Measurements

An induction motor of 35 kW is fed from a Vacon inverter. A DC generator is the motor load. The switching frequency of the inverter is fixed at 3 kHz. Five acceleration vibration sensors are placed in different parts of the motor. Three of them are placed in the cooling surface of the motor frame. One is placed in the covering surface of the frontal bearing and the other one near the cooling fan in the back part of the motor.

The signals given by the vibration sensors are amplified through charge amplifiers Bruel & Kjaer 2635. The amplified signals are the transient recorder inputs. The recorded transient is calibrated using a true root mean square voltmeter connected in the amplifier output. This calibration is made in such a way that the recorded measurements in the transient recorder are in acceleration unity. In Fig. 2, the measurement set-up is presented.

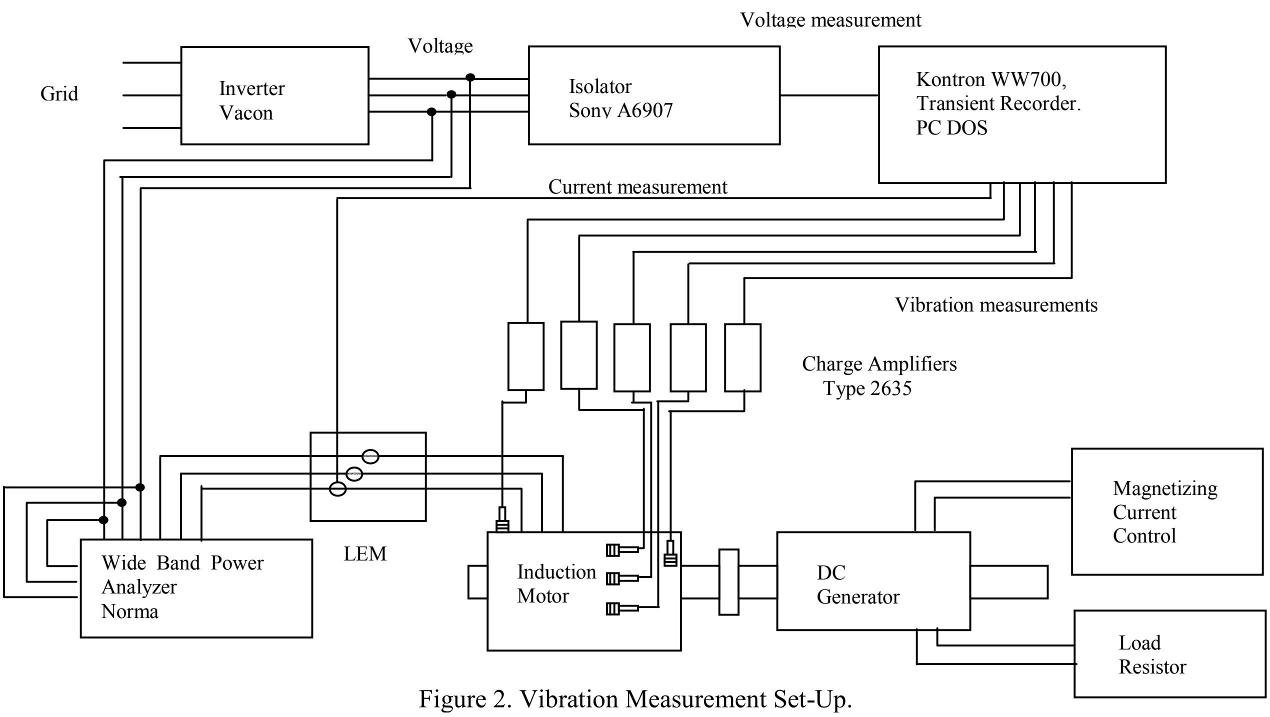

Figure 2. Vibration Measurement Set-Up.

Measurements are carried out with motor in healthy condition and with motor under three rotor fault situations: one broken rotor bar, two broken rotor bars and three broken rotor bars plus end ring broken. The sampling frequency is 40 kHz and the number of samples in a data set is 20 thousands. Three load situations are considered: no load, half load and full load.

In Fig. 3, vibration measurements in a healthy situation with no load are presented. It can be seen that one of the vibration measurements is considerable higher than measurements from the other sensors. The same phenomenon can be seen from Fig. 4, where measurements in a broken bar situation with no load are presented. Further in Fig. 5, measurements in two broken bar situation with no load are presented. Again one of the sensors produces the most relevant vibration signal, but it is surprising than in this more serious fault situation all vibrations are generally lower than in one broken bar situation.

4.2 Broken Rotor Bar Detection with ICA and SVM

FastICA algorithm is used to calculate IC's of vibration measurements. Before applying FastICA, the vibration measurements are whitened with PCA.

In Fig. 6, Welch's PSD estimates of the first IC in healthy and faulty situation are presented in the same picture (Hanning window sized 500 samples with 250 overlapping samples). One can see that they differ from each other, even if the vibration signals seem to be quite similar in time domain.

LS-SVM based classifier is built to discriminate between the healthy and broken rotor bar condition. It is tested also with measurements from two broken bar situation and three broken bar and broken end ring situation. LS-SVM has a kernel function that corresponds to the first order polynomial and the upper bound for Lagrange multipliers was chosen to be equal to 10.

Training data is formed by calculating Welch's PSD estimates from different parts of the first IC of vibrations in the healthy and broken rotor bar situations. In earlier figures, only no load situations are plotted, but we will take into account also two other load situations in training: half load and full load. The training data consists of 2x3x15 = 90 samples so that from both motor conditions and from all load situations there exist 15 samples. The other faults than one broken rotor bar fault are used only in testing the classifier, so that in total there are 240 samples for testing: 45 from healthy situation,

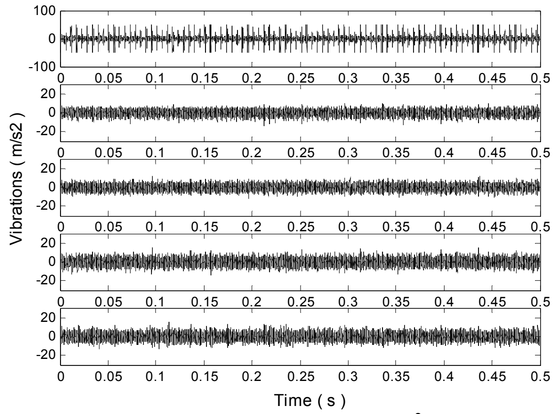

Figure 3. Vibration Measurements (m/s^2) in a Healthy Situation.

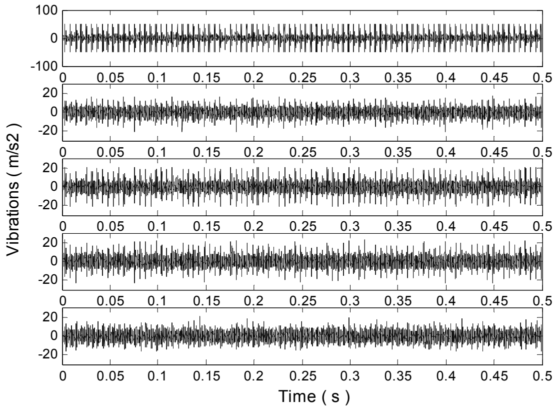

Figure 4. Vibration Measurements (m/s^2) in a Broken Bar Situation.

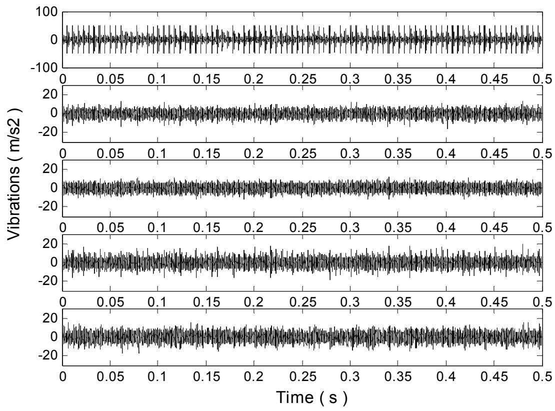

Figure 5. Vibration Measurements (m/s^2) in a Two Broken Bar Situation.

45 from one broken rotor bar situation and 90 from two broken rotor bars and 60 from three broken rotor bars and broken end ring. In the last case, full load measurements were not done. All of the test samples are correctly classified.

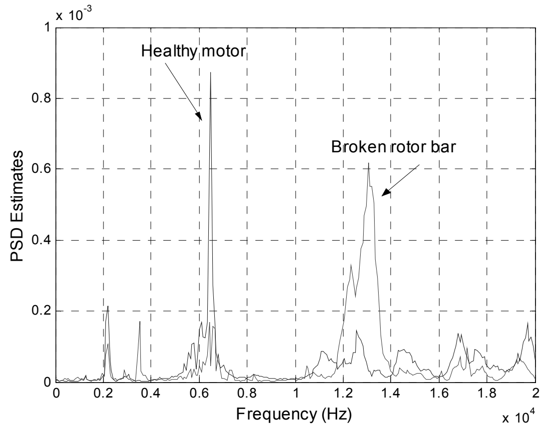

Figure 6. Welch's PSD estimates of the First IC of Vibrations in Healthy Situation and in Broken Rotor Bar Situation with No Load.

For comparison, a classifier that uses the main vibration component as a fault indicator instead of the 1st IC of all vibrations was also trained. Training and testing data sets were formed in a same way as earlier. Now 5% of test samples were wrongly classified. All of these samples were two broken rotor bar samples that were wrongly classified to be healthy.

Further, a classifier that uses the first principal component (PC) of vibrations as fault indicator was trained. Again, all the test samples were correctly classified. Even if the difference between using pure vibrations or IC's or PC's of vibrations in fault detection is quite small, this could indicate that faults are more easily detected from fused vibration measurements than pure vibrations. IC's and PC's contain more information on all vibration measurements than any of the individual vibration measurements. Using all vibration components for classification might improve the results without any data fusion, but at the same time computation would become heavier.

5. Conclusion

ICA was applied to multi-channel vibration measurements of an induction motor to fuse the measurement information of several channels, and provide robust and reliable broken rotor bar detection. IC's were found from the measurement set with a FastICA algorithm and their PSD estimates were calculated with Welch's method. A SVM based classification routine was applied to the PSD samples to perform the fault diagnosis. Similar classification was applied directly to PSD estimates of vibration measurements and to PSD's of PC's of the measurements. Results gained with the two data fusion methods were compared to the results gained without data fusion. Both methods enhance the fault diagnostics routine.

In this case, both data fusion methods, ICA and PCA, resulted in equally excellent performance of broken rotor bar detection. However, further studies are required with measurements from other faults to conclude overall usefulness of data fusion of vibration measurements, because the broken rotor bar was quite easily detected also based on the main vibration measurement. Also, building a multi-class classifier for detection of several faults may degrade the classification results.

References

[1] G. Betta, C. Liguori, A. Paolillo, A. Pietrosanto, A., A DSP-based FFT analyzer for the fault diagnosis of rotating machine based on vibration analysis, *Proc. IEEE Conf. on Instrumentation and Measurement Technology*, Budabest, Hungary, 2001, 572-577.

[2] R.F.M. Marcal, M. Negreiros, A.A. Susin, J.L. Kovaleski, Detecting faults in rotating machines, *IEEE Instrumentation & Measurement Magazine*, *3*(4), 2000, 24-26.

[3] P.A. Laggan, Vibration monitoring,, *Proc. IEE Colloquium on Understanding your Condition Monitoring*, 1999, 1-11.

[4] A. Ypma, P. Pajunen, P, Rotating machine vibration analysis with second-order independent component analysis, *Proc. of the Workshop on ICA and Signal Separation* , Aussois, France, 1999, 37-42.

[5] A. Ypma, D.M.J. Tax, R.P.W Duin, Robust machine fault detection with independent component analysis and support vector data description, *Proc. IEEE Signal Processing Society Workshop on Neural Networks for Signal Processing*, Madison, Wisconsin, USA, 1999, 67–76.

[6] N. Cristianini, J. Shawe-Taylor, *Support vector machines and other kernel-based learning methods* (Cambridge University Press, 2000).

[7] A. Ypma, A. Leshem, Blind separation of machine vibration with bilinear forms, *Proc. 2nd Int. Workshop on ICA and Signal Separation,* Helsinki, Finland, 2000, 405-410.

[8] A. Ypma, A., Leshem, R.P.W. Duin, Blind separation of rotating machine sources: bilinear forms and convolutive mixtures*, Neurocomputing - Special Issue on ICA/BSS*, *49*(1-4), 2002, 349-368.

[9] G. Gelle, M. Colas, C. Serviere, Blind source separation: A tool for rotating machine monitoring by vibration analysis, *Journal of Sound and Vibration*, *248*(5), 2001, 865-885.

[10] A. Hyvärinen, J. Karhunen, E. Oja, *Independent component analysis* (John Wiley & Sons, Inc., NY 2001).

[11] A. Hyvärinen, Fast and robust fixed-point algorithms for independent component analysis, *IEEE Trans. on Neural Networks, 10*(3), 1999, 626-634.

[12] J.A.K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, J. Vandewalle, *Least squares support vector machines*, (World Scientific, Singapore, 2002).