Characterization of the mechanical vibration signals associated with unbalance and misalignment in rotating machines, using the cepstrum transformation and the principal component analysis.

C L Sandoval-Rodriguez[1,] J G Ascanio Villabona [1], C G Cárdenas-Arias[1], A D Rincon-Quintero[1-2], B E Tarazona-Romero[1].

[1] Energy Systems, Automation and Control Research Group (GISEAC), Design and materials research group (DIMAT), Electromechanical Engineering, Faculty of Natural Sciences and Engineering, Technological Units of Santander (UTS), Road No 9-82 students, 680,005, Bucaramanga, Colombia.

[2] Energy in Building Research Group (ENEDI), Doctoral Program in Energy Efficiency and Sustainability in Engineering and Architecture, Department of Machines and Thermal Engines, University of the Basque Country (UPV / EHU), Engineer Torres Quevedo Square, 1, 48013, Bilbao, BI, Spain.

candoval@correo.uts.edu.co

Abstract. In the present document, the Cepstrum transform and the analysis of principal components were used to differentiate amplitudes in the mechanical vibrations produced by unbalance and misalignment with respect to a reference group. This document requires three stages. It begins with levelling in order to establish the control group. The unbalancing was carried out with a known mass located in the two radial distances of the first and second flywheels. The misalignment was made by running the sliding supports back 0.5, 1.0 and 1.5 degrees. In the second stage, Matlab algorithms were created for both cepsctrum and main component analysis. In the last stage the obtained data were analyzed identifying the differences that may exist in the analyzed records. The project focused on the use of Matlab to find differences at a frequency of 30 Hz. The results obtained made it possible to determine that it is possible to find differences with the proposed methodology.

1. Introduction.

Rotating machines have different operating states during its lifetime. These may be normal type or operation failures, which may be due to little study of the machine, and critical operation, which can affect permanently to different parts making up the machine [1]. Because of this industry began to adopt different types of maintenance on rotating machines in order to address the failures that may occur in these teams; preventively before the failure, correctively occurs when the failure occurs and is investigated because occurred; and fails when it seeks to replace damaged parts [1].

In rotary machines 2 important affectations studied in this project are presented. The first is due to improper mounting of the machine by poor manufacture of the bed or poor mechanical coupling between drive and driven shafts; this involvement is known as alignment fault [2].

On the other hand, the second involvement is concerned with wear, incorrect selection of a failed component, faulty manufacture of the components of the machine, some dynamic fracture machine part, among other causes; by which it can be produced by unbalance failure [3]. Although the type of failure can be detected in time, other inherent flaws in rotating machinery problems arise, which have to do with other factors such as; interpretation of curves, judgment and subjectivity of the analyst. The interpretation of the spectra in vibration analysis equipment or specialized software depends on the experience, knowledge and judgment on the analyst applied vibration; to determine whether visible sometimes audible failures are due to imbalance, misalignment, mechanical looseness, worn bearings, bent arrows or electrical problems [2]. In this sense, the solution to this problem focuses on the possibility of finding a method to moderate the interpretation of the information from the vibration, the analyst's subjectivity; and also serve to support maintenance professionals for a correct decision in determining faults in rotating machines.

Importantly, the scientific and academic community has been concerned about this problem and have generated around this, various alternatives both in terms of feature extraction [4-7], from mechanical vibrations, as in systems that allow discrimination of each operating mode [8-9] and focused data compression for the application of classification and troubleshooting rotating machinery [10]. In this regard, it has raised for this work to extract features related to the behavior of a rotating system affected by unbalance and misalignment independently.

The methodology used in the proposed solution is that the descriptive desired characterize faults; correlational because a reference group is set to analyze faults unbalance and misalignment in rotating machinery [11]. Similarly, the method used is the deductive because based on overall theories transformed cepstrum and principal component analysis in solving particular problems as proposed in this work [12]; on the other hand, the technique used to validate all results outlined in this research is the experimental [11] type.

Matlab is used to analyze three general types of samples; The control group, the group of imbalance and misalignment group. In Matlab different algorithms in order to select samples from the same variable that are grouped in reference to the amplitude at the fundamental frequency created. It also created algorithms to calculate the cepstrum transformed to perform principal component analysis 8 observations, to obtain the Euclidean distance and different characteristics in the analysis of 6 variables statistics.

2. Cepstrum Analysis and Principal Component Analysis

The transformed cepstrum Oppenheim and Schafer as [13], with Bogert emerged in 1963, referring to a way to analyze signals in the time domain, which did not belong to the frequency domain; cepstrum can be defined as the spectrum of the natural logarithm of the spectrum as shown in equation 1.

$$X[n] = IFFT\{log|FFT(x[n])|\} \quad (1)$$

In this equation the expression IFFT corresponds to the discrete inverse Fourier transform, the expression FFT is the discrete Fourier transform, the expression log is the logarithm base 10 and the expression x [n] is the signal in the time domain. With these coefficients aspects of the spectra such as the envelope or detailed characteristics [14] recovered.

It is clear that the drawback that occurs in the technique of feature extraction by applying transformations of the signal is the high dimensionality of the resulting space, which would form very large vectors description. Given this, the dimensionality reduction techniques help to preserve the most important information while the dimension of the transformed space is reduced [15-16]. The idea of the principal component analysis is to estimate the variability of each data set and retain greater variance components according to certain criteria [17] in order to use them as representative characteristics of each mode.

The first principal component is the linear combination of coefficients with a high variability; as shown in Equation 1. The second is due to the linear combination of the coefficients of the second greater variability and so on.

$$Xp_n = \sum Cp_i * \lambda p_i \qquad (2)$$

Where Xp is the main component, the subscript n indicates the order of the component, Cp is the coefficient related to the variance of the coefficient λ. After calculating the principal component s are retained only those showing greater variability using different criteria. For this work four criteria were implemented:

Jolife criteria: Only the components whose variance is greater than 0.7 are retained.

Kaiser criterion; Are retained components whose variance is greater than average.

standard criteria: The components are retained whose variance is below 2% (Criterion elbow)

Cumulative variance criterion: The components are retained whose variance is greater than 1% of the total cumulative variance of all components. With the application of the different criteria the components to be applied to the algorithm that allow differentiation are retained.

3. Sampling and observations

The calculation of the samples was conducted using Equ (3) in infinite sample size populations without known antecedent.; where Z is the confidence of the sample, P is the probability of success, Q is the probability of failure and E is the allowable error.

$$n = \frac{ZPQ}{E} \qquad (3)$$

For a population in the conditions described should be taken Z = 1.6448, the probability of success and failure should equal P = Q = 0.5, and the allowable error E = 0.01 was taken.

$$n = \frac{1.6448 * 0.5 * 0.5}{0.01} = 67.63 \cong 68$$

In Table 1, the variables used for performing bank failures in the vibration according to the limitations found in preliminary tests are described.

Table 1. Description of variables.

VARIABLE	DESCRIPTION
GCAB	Control group alignment and balancing
DA05G	Movement of the support for 0.5 ° degrees
DA10G	Movement supports to 1.0 ° degrees
DA15G	Movement supports to 1.5 ° degrees
DBV1R1	Place a mass m1 in the wheel 1 to the distance r1
DBV1R2	Place a mass m1 in the wheel 1 to the distance r2
DBV2R1	Place a mass m1 in the wheel 2 to the distance r1
DBV2R2	Place a mass m1 in the wheel 2 to the distance r2

Point for data collection, is located at the farthest support the engine in horizontal position to half the distance of the bearing and the signal acquisition was performed using the piezoelectric accelerometer Dytran. All samples were performed at 30 Hz, due to the limitations found in preliminary tests.

Installation and data acquisition was performed by means of a sensor Dytran brand, acquisition card National Instruments and Labview program [18]. In Fig.1 the hardware which was used for the tests shown.

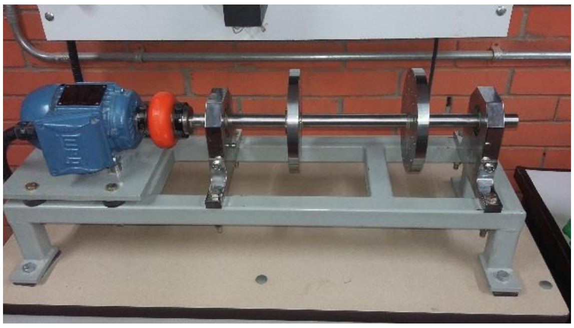

Figure 1. Mechanical Hardware Test.

3.1. *Obtaining the characteristics using cepstrum coefficients into mel scale and principal components analysis.*

For comparison of characteristics comparison algorithm using Euclidean distance used by [19-20] it was used. This system is based on a priori knowledge base characterizing cases the required class. a comparison control group compared to the other seven groups shown in Table 2.

Table 2. Euclidean distance of CCM related variables

Variables Name	Relationship	no CCM	DE * e-3
CEPDA1	GCABvsDA05g	1	3.93322
CEPDA2	GCABvsDA10g	1	4.82292
CEPDA3	GCABvsDA15g	3	10.94881
CEPD1	GCABvsDBv1r1	4	11.80011
CEPD2	GCABvsDBv1r2	3	16.63780
CEPD3	GCABvsDBv2r1	2	17.28532
CEPD4	GCABvsDBv2r2	2	32.77406
DE: Euclidean distance.			
CCM: Cepstrum coefficient on the Mel scale.			

It should be noted that for the principal component analysis (PCA), the Fourier spectrum is calculated and spectral coefficients, applying the different criteria the components which in turn are the characteristics of each observation are retained. In this regard the components shown in Table 3 were obtained.

Table 3. Components retained on each criterion

Criterion	Retained components	GROUPS OBSERVED
Jolife	7	everyone
Kaiser	2	everyone
Standard	1	everyone
Cumulative variance	2	everyone

It is important to note that not necessarily the criterion to retain the minimum number of components will best performance features. To probe the above was necessary to apply the algorithm developed by [20] to give the results shown in Table 4 (case-based reasoning-RBC).

Table 4. Percentages classification and standard deviations for each criterion.

Criterion	PERCENTAGE OF CLASSIFICATION	Standard Deviations
Jolife	80.2	2
Kaiser	78.66	4,21
Standard	82	6,32
Cumulative variance	80.2	6.96

According to the above criteria Jolife (good classification rate and high concentration of data), for characterization as described in Table 5 was made.

Table 5. Statistical employees

Statistical	Calculation method	Description
E1: Average value	$T1 = \frac{\sum_{n=1}^{N} x(n)}{N}$	The average value in a set of data within a reasonable range represents a trend of values to a representative value for all [21]
E2: Standard Deviation	$T2 = \sqrt{\frac{\sum_{n=1}^{N} (x(n) - T1)^2}{N-1}}$	The standard deviation is the root of the sum of the squared difference values of the signal with the average value between the total data minus 1. The measure indicates how scattered the data are the mean value zero being an ideal value where the data signal and its average value agree [21]
E3: RMS root mean square	$T3 = \sqrt{\frac{\sum_{n=1}^{N} (x(n))^2}{N}}$	The root mean square or RMS is the root of the sum of the square of a number by the total amount of data. By relying heavily on the average value of the RMS data it is also susceptible to extreme outliers [21]
CV: Coefficient of variation	$CV = \frac{T2}{T1}$	the coefficient of variation which is defined as the standard deviation about the mean value of the signal [21]

3.2. Statistical results calculated

After applying statistical both mel cepstral coefficients in the scale and the main components for the criterion Jolife the results shown in Table 6. (Note that the vectors have compared the cepstral coefficients and the principal components or for each group were obtained observation)

Table 6. Results of the characterizations for the cepstral coefficients of Table 2.

Variable	statisticians			
	E1 * e4	E2 * e-4	CV%	E3 * e-4
GCAB	51.9257	3.0682	5.91	52.0149
DA05g	52.2840	3.2627	6.24	52.3842
DA10g	52.9456	4.3471	8.21	53.1211

DA15g	38.4259	3.6330	9.45	38.5947
DBv1r1m	62.8185	12.5058	19.91	64.0333
DBv1r2m	69.4527	2.8814	4.15	69.5116
DBv2r1m	77.7885	3.7489	4.82	77.8775
DBv2r2m	100.5636	5.0916	5.06	100.6905

E1: Average value
E2: Standard Deviation
E3: RMS root mean square
CV: Coefficient of variation

From the above it can be seen that the average value to differentiate each of the variables in Table 1, where the variables for the unaligned group are above and below the reference group and variables unbalanced group are above reference group. Variable statistical standard deviation E2 does not provide much information al1, but if the coefficient of variation is used, it can be concluded that the overall variables in Table 5, do not exceed 10% coefficient of variation; however DBv1r1 variable passes this value and is located in a coefficient of variation of approximately 20%; so in general all the variables do not have a too large standard deviation.

The variable statistic root mean square E3 allows differentiating each of the variables in Table 6, where the variables for the unaligned group are above and below the reference group and variables unbalanced group are above group reference.

4. Conclusions

The anomalies in rotary machines can be identified up to the fourth cepstrum coefficient in mel scale. The Euclidean distance made it possible to find differences between the Mel scale cepstrum coefficients for unbalance and misalignment conditions in rotary machines. On the other hand, the principal components that represent greater discrimination power were calculated taking into account the Jolife criterion, obtaining characteristics such as the mean value, standard deviation, variation coefficient and quadratic mean root. Additionally, the coefficient of variation allowed to determine that the variables studied for the unbalance and misalignment conditions do not have a very high standard deviation. Finally, it was evidenced that both the mean value and the root mean quadratic allow to differentiate all the groups when composing a characteristic vector containing the cepstrales coefficients in mel scale and the principal components obtained from the Fourier coefficients of the spectrum.

References

[1] Pérez González A, Rodríguez Cervantes P J, & Sancho Brú J L, (2007). Mantenimiento mecánico de máquinas. Universitat Jaume I, Castelló de la Plana, España..

[2] Gómez de León F C. (1998). Tecnología del mantenimiento industrial. Murcia: EDITUM.

[3] Rao S S, (2012). Vibraciones Mecánicas (5a edición) Pearson Educación., pp. 651-658.

[4] Li C, Zhang W E I, Peng G, & Liu S (2017). Bearing fault diagnosis using fully-connected winner-take-all autoencoder. *Access IEEE, 6, 6103-6115.*

[5] Feng Z, Zhang D, & Zuo MJ (2017). Adaptive mode decomposition methods and their applications in signal analysis for fault diagnosis machinery: a Review with examples. *IEEE access, 5, 24301-24331.*

[6] Wang D, Tsui KL, & Miao Q. (2017). Prognostics and health management: A review of bearing and gear vibration based health indicators. *Access IEEE, 6, 665-676.*

[7] Lu Y, Li Q, Pan Z, & Liang, SY (2018). Prognosis of bearing degradation factor using gradient Variable forgetting RLS combined time series model With. *Access IEEE, 6, 10986-10995.*

[8] Duan X, Zhao T, Li T, Liu J, Zou L, & Zhang L. (2017). Method for diagnosis of on-load tap changer based on wavelet theory and support vector machine. *The Journal of Engineering, 2017 (13)*, **2193-2197.**
[9] Liu S, Sun Y, & Zhang L. (2018). A novel fault diagnosis method based on noise-assisted DMEM and neural network for functional fuzzy rolling element bearings. *Access IEEE, 6*, 27048-27068.
[10] Ahmed H, & Nandi AK (2018). Compressive sampling and feature ranking framework for classification With bearing fault vibration signals. *Access IEEE, 6*, 44731-44746.
[11] Hernández Sampieri R, Fernández Collado C & Baptista Lucio P. (2010). Metodología de la investigación. pp. 102-160. 2006.
[12] Bernal, C. A. (2006). Metodología de la investigación (No. 001.42 B4564m Ej. 1 022619). Pearson, **pp. 56**. [Online]. Available: https://books.google.com.co/books?id=h4X_eFai59oC&pg=PA56&dq=metodo+deductivo& hl=es-419&sa= X & SQI = 2 & pjf = 1 & ved = 0ahUKEwiNtKKFzYTUAhVDfiYKHUtAA10Q6AEIITAA # v = 1page & q = 20% deductive method & f = false.
[13] Oppenheim, A V & Schafer R W. (2004). From frequency to quefrency: A history of the cepstrum. *IEEE signal processing Magazine*, **21(5), 95-106**. [Online]. Available in: https://www.researchgate.net/ profile / Av_Oppenheim / publication / 3321562_From_Frequency_to_Quefrency_A_History_of_the_Cepstrum / links / 5519617d0cf2d241f3564c58.pdf
[14] Quiroga J., G. Trujillo, S. and Quintero, "Study of incipient bearing failure using the technique of the envelope and cepstrum" *Ingeniare, vol. 20*, no. 3, **pp. 350-359**. Nov. 2012 [Online]. Available in:http://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0718-33052012000300009&lng=
[15] Smith LI (2002). A tutorial on principal components analysis.
[16] Rao CR (1964). The use and interpretation of principal component analysis in applied research. Sankhyā: The Indian Journal of Statistics, Series A, 329-358.
[17] PEÑA D. Multivariate analysis data, Chapter 5.McGraw-Hill. ISBN: 8448136101.
[18] Forero W, Lizcano D, & Col11 J. "Implementation of a bank for the analysis of mechanical vibrations generated by unbalance and misalignment in rotating machines," Undergraduate Thesis, Technological Units of Santander, Bucaramanga, Colombia, 2016.
[19] Camacho Navarro, J; "Numerical adaptation and validation expert for structural health monitoring through pattern recognition system". Industrial University of Santander, Bucaramanga 2010.
[20] Sandoval Rodriguez, C L. Characterization of the dynamics of basic hand movements from the electromyography activity forearm. Industrial University of Santander 2013.
[21] Martinez C, statistics and sampling. 13th Edition. ECOE Editions, pp. 303-310. 2012