

# Robust Industrial Machine Sounds Identification Based on Frequency Spectrum Analysis

Antoni Grau, Yolanda Bolea, and Manuel Manzanares

Automatic Control Dept, Technical University of Catalonia UPC, Barcelona, Spain  
{antoni.grau,yolanda.bolea,manolo.manzanares}@upc.edu

**Abstract.** In order to discriminate and identify different industrial machine sounds corrupted with heavy non-stationary and non-Gaussian perturbations (high noise, speech, etc.), a new methodology is proposed in this article. From every sound signal a set of features is extracted based on its denoised frequency spectrum using Morlet wavelet transformation (CWT), and the distance between feature vectors is used to identify the signals and their noisy versions. This methodology has been tested with real sounds, and it has been validated with corrupted sounds with very low signal-noise ratio (SNR) values, demonstrating the method's robustness.

**Keywords:** wavelets, Fast Fourier Transformation, non-speech sound.

## 1 Introduction

A common problem encountered in industrial environments is that the electric machine sounds are often corrupted by non-stationary and non-Gaussian interferences such as speech signals, environmental noise, background noise, etc. Consequently, pure machine sounds may be difficult to identify using conventional frequency domain analysis techniques as Fourier transform [1], and statistical techniques such as Independent Component Analysis (ICA) [2]. It is generally difficult to extract hidden features from the data measured using conventional spectral techniques because of the weak amplitude and short duration of structural electric machine signals, and very often the feature sound of the machine is immersed in heavy perturbations producing hard changes in the original sound. For these reasons, the wavelet transform has attracted increasing attention in recent years for its ability in signal features extraction [3][4], and noise elimination [5]. While in many mechanical dynamic signals, such as the acoustical signals of an engine, Donoho's method seems rather ineffective, the reason for their inefficiency is that the feature of the mechanical signals is not considered. Therefore, when the idea of Donoho's method and the sound feature are combined, and a de-noising method based on the Morlet wavelet is added, this methodology becomes very effective when applied to an engine sound detection [6].

In this work, we propose a new approach in order to identify different industrial machine sounds, which can be affected by non-stationary noise sources. This paper is organized as follows. In Section 2 the proposed methodology is overviewed. Next the

Morlet wavelet transform for denoising the acoustical signals is explained as well as the feature extraction procedure. Some interesting experimental results are presented.

## 2 Methodology

In this section the proposed methodology to identify machine sounds highly corrupted with non-stationary and non-Gaussian perturbations is presented. The original sounds corrupted with noise will be denoted as “noisy” sounds. The procedure consists of the following steps:

- 1) Given a certain number of “original” sounds recorded from different machines, they will be heavily corrupted with different non-stationary non-Gaussian noises giving different SNR (signal-noise ratio). The problem is that in the temporal space it is impossible to distinguish between two corrupted signals. Due to this reason, we propose the step 2.
- 2) We will need to work in the frequency space. In order to overcome the lack of identification among noisy sounds, a denoised version is proposed in this article. To carry out the purification (denoising) process, the Morlet wavelet transform (Section 3) will be used as a filter step. A reconstructed version of the noisy signal is generated after a filtering operation, setting to zero some of the wavelet coefficients. Therefore, the frequency spectrum of this denoised signal will be used as a source of the feature extraction process.
- 3) The features extracted from the frequency spectrum are the basis for identifying the original sounds. This procedure will be done by calculating the distance between the feature vectors extracted from the original signal and the denoised one.
- 4) To evaluate the robustness of our methodology, we proceed reducing the SNR level (increasing the level of noise) until the distance between the original and its denoised signal exceeds two values: first, the value of the distance between the original signal and its noisy version ( $d(s_i, s_{i\_denoised}) > d(s_i, s_{i\_noisy})$ ), and second, the value of the distance between the original signal and a denoised signal coming from another original signal ( $d(s_i, s_{i\_denoised}) > d(s_i, s_{j\_denoised})$ ), being  $s_i$  and  $s_j$  two different original signals. This fact demonstrates the robustness of the proposed methodology because high levels of noise can be added without misidentifying the signals.

## 3 Wavelet Transform and Feature Extraction

### 3.1 Review of Wavelet Transform

The wavelet was originally introduced by Goupilland et al. in 1984 [7]. Let  $\psi(t)$  be the basic wavelet function or the mother wavelet, then the corresponding family of daughter wavelets consists of

$$\psi_{a,b}(t) = |a|^{-1/2} \psi\left(\frac{t-b}{a}\right) \quad (1)$$

where  $a$  is the scale factor and  $b$  the time location, and the factor  $|a|^{-1/2}$  is used to ensure energy preservation.

The wavelet transform of signal  $x(t)$  is defined as the inner product in the Hilbert space of the  $L^2$  norm, as shown in the following equation

$$W(a, b) = \langle \psi_{a,b}(t), x(t) \rangle = |a|^{-1/2} \int x(t) \psi_{a,b}^* dt \quad (2)$$

Here the asterisk stands for complex conjugate. Time parameter  $b$  and scale parameter  $a$  vary continuously, so that transform defined by Eq. (2) is also called a continuous wavelet transform, or CWT. The wavelet transform coefficients  $W(a, b)$  can be considered as functions of translation  $b$  for each fixed scale  $a$ , which give the information of  $x(t)$  at different levels of resolution. The wavelet coefficients  $W(a, b)$  also measure the similarity between the signal  $x(t)$  and each daughter wavelet  $\psi_{a,b}(t)$ . This implies that wavelets can be used for feature discovery if the wavelet used is close enough to the feature components hidden in the signal.

For many mechanical acoustic signals impulse components often correspond to the feature sound. Thus, the basic wavelet used for feature extraction should be similar to an impulse. The Morlet wavelet is such a wavelet defined as

$$\psi(t) = \exp(-\beta^2 t^2 / 2) \cos(\pi t) \quad (3)$$

### 3.2 Feature Extraction Using the Morlet Wavelet

The most popular algorithm of wavelet transform is the Mallat algorithm. Though this algorithm can save a lot of computations, it demands that the basic wavelet is orthogonal. The Morlet wavelet is not orthogonal. Thus, the wavelet transform of the Morlet wavelet has to be computed by the original definition, as shown in Eq. (2). Although the CWT brings about redundancy in the representation of the signal (a one-dimensional signal is mapped to a two-dimensional signal), it provides the possibility of reconstructing a signal. A simple inverse way is to use the Morlet's formula, which only requires a single integration. The formula is:

$$x(t) = C_{\psi}^{-1} \int W(a, b) \frac{da}{a^{3/2}} \quad (4)$$

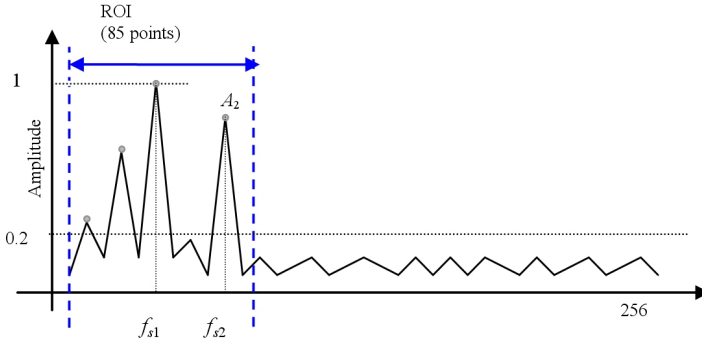
where

$$C_{1\psi} = \int_{-\infty}^{\infty} \hat{\psi}^*(\omega) / |\omega| d\omega \quad (5)$$

It is valid when  $x(t)$  is real and either  $\psi(t)$  is analytic or  $\hat{\psi}(\omega)$  is real. The condition is satisfied by the Morlet wavelet. If the wavelet coefficients  $W(a, b)$ , corresponding to feature components, could be acquired, we could obtain the feature components just by reconstructing these coefficients. In calculations, the feature coefficients should be reserved and the irrelevant ones set to zero, then the signal can be denoised by using formula Eq. (4). Thus, the key to obtaining the denoised signal is how to obtain these feature coefficients. A threshold  $T_w$  should be set in advance, but it is not evident to choose it properly. This threshold indicates the value from which wavelet coefficients

must be set to zero. The basic rule for threshold choice is that the higher the correlation between the random variables, the larger the threshold; and the higher the signal-noise ratio (SNR), the lower the threshold. In practice, the choice of the threshold  $T_w$  mainly depends on experience and knowledge about the signal. In fact, the quantitative relation between the threshold  $T_w$  and the SNR still remains an open question.

Since the sound signal is reconstructed with the modified wavelet coefficients, the frequency spectrum is calculated using the Fast Fourier Transform. In the results, it is easy to appreciate the good level of denoising that wavelet transform is yielding.



**Fig. 1.** Definition of features upon the frequency spectrum

Observing in detail such a spectrum, machine sounds have the characteristic of being formed by  $m$  harmonics with different amplitude and located at different frequencies. Therefore, these signals are divided in regions-of-interest (ROI) consisting of 85 points (over 256-point FFT). The width of the ROI is chosen so that this area is containing information enough to do the analysis and, consequently, the main fundamental harmonics in the signals can be found inside the ROI. A normalization step is done in the signal spectrums (between  $[0...1]$ ). Due to the properties of mechanical acoustic signals, the most important parameters that we consider to characterize the machine sounds are: the frequency of the first harmonic ( $f_{s1}$ ), the amplitude of the second harmonic ( $A_2$ ) and the number of significant harmonics (with an amplitude  $> 0.2$ ), note that the first harmonic amplitude will be always 1, that is, the maximal value of normalization. For every sound signal these three features will form the feature vector. In Fig. 1 the features can be observed over a generic spectrum. Nevertheless, other features have been tried such as the frequency of the second harmonic ( $f_{s2}$ ) but the results do not improve.

### 3.3 Distance Between Feature Vectors

A distance measure is a function that associates a non-negative numeric value with (a pair of) sequences, with the idea that a short distance means greater similarity. Distance measures usually satisfy the mathematical axioms of a metric. In the Euclidean space  $\mathbb{R}^n$ , the distance between two points is usually given by the Euclidean distance (2-norm distance). Other distances, based on other norms, are sometimes used instead. In this article we use the Minkowski distance of order 2 (2-norm distance) defined as:

$$d(x, y) = \sqrt{\sum_{i=1}^n |x_i - y_i|^2} \quad (6)$$

for a point  $(x_1, x_2, \dots, x_n)$  and a point  $(y_1, y_2, \dots, y_n)$ . Other norms have been tested but the results do not improve, because the distance measure is not relevant. In our opinion, after the tests, the most relevant in the identification process is the feature extraction process. The goodness of the identification depends on the denoising methodology for this kind of signals (machine sounds) and the selection of the most representative features from the spectrum.

For identifying the noisy signals and, effectively, to know what original signal machine sounds they come from, this distance will serve as a dissimilarity measure, defining:

**Table 1.** Different distances between signals

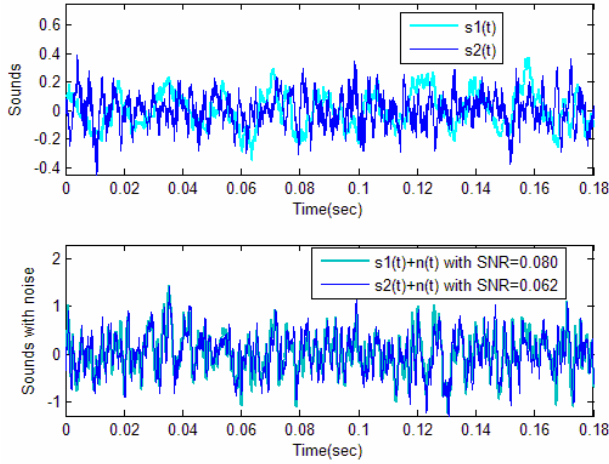
Acronym	Definition	Description
<i>dsii</i>	$d(s_i, s_{i\_noisy})$	Distance between an original sound and one of its noisy versions
<i>dsij</i>	$d(s_i, s_{j\_noisy})$	Distance between an original sound and one of a noisy version of another original signal. Cross distances.
<i>dsii_denoised</i>	$d(s_i, s_{i\_denoised})$	Distance between an original sound and one of its denoised versions
<i>dsij_denoised</i>	$d(s_i, s_{j\_denoised})$	Distance between an original sound and one of a denoised version of another original signal. Cross distances.

## 4 Experimental Results

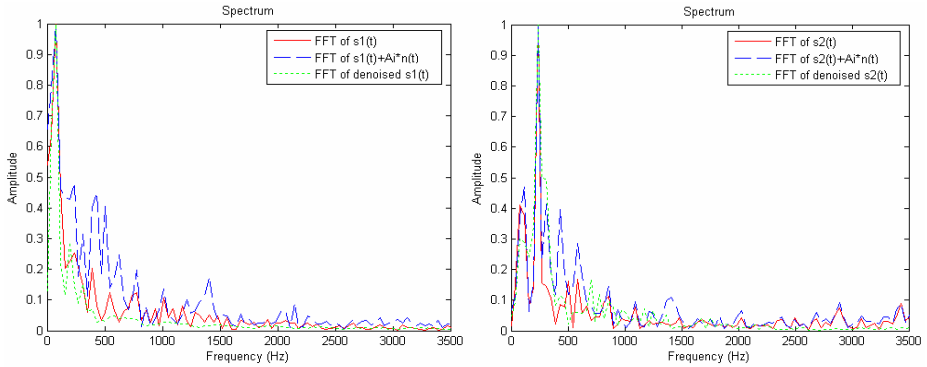
In our experiments, we will demonstrate the identification capacity and the robustness of our methodology when the input signals are highly corrupted with low levels of SNR (high levels of noise).

Initially, we use two original sound signals recorded in real industrial environments:  $s_1$ , a car factory assembly line sound; and  $s_2$ , a ship engine room sound. These two signals have been corrupted with a non-stationary, non-Gaussian and colored noise, called babble noise ( $n$ ), with different amplitudes, creating a set of 125 noisy signals with decreasing SNR, ranging from 12 to 0.5. The babble noise [8] represents more than one hundred people speaking in a canteen. All these sounds have been recorded at a frequency sample of 19.98 kHz, 16-bit, mono. Fig. 2 (up) shows original sounds and Fig. 2 (down) their corresponding corrupted sounds before the denoising process. It is important to remark the difficulty to discriminating between both noisy signals.

In the denoising process, the threshold  $T_w$  is set to 30% of the maximum wavelet coefficients of the original signals, fulfilling the basic rule stated in that Section 3.2. Fig. 3 shows the similitude among the different spectra for the same signal when it is completely clean (original), when it is completely noisy and finally when it has been denoised after the wavelet reconstruction. It can be observed that denoised spectrum is similar to the clean one, and therefore, the features are quite similar compared with the features obtained from the noisy signal.



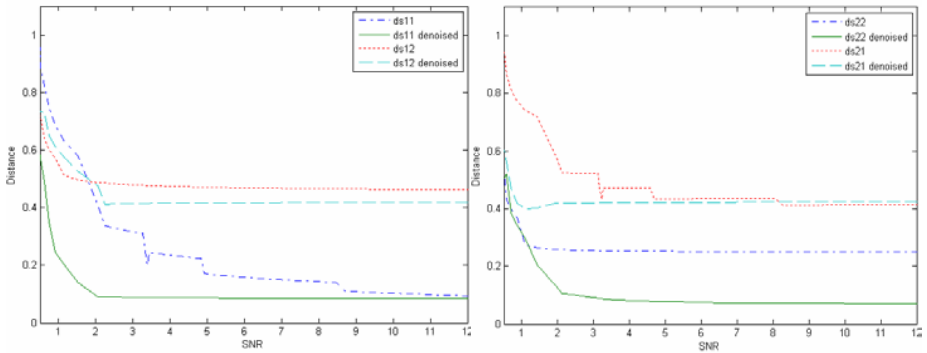
**Fig. 2.** (Up) Machine sounds; (down) with babble noise



**Fig. 3.** Spectra of original, noisy and denoised sounds of  $s_1$  (left) and  $s_2$  (right)

Fig. 4 shows the distance between different signals versus the SNR. It can be observed that for both signals ( $s_1$  and  $s_2$ ) the distance from each original signal to its denoised versions is always lower than the distance between each original signal to its noisy versions, with a wide range of low SNR values. This fact is very important because it demonstrates that the denoising process works very well, reducing the added noise without losing the implicit frequency information of original signal spectrum.

Another consequence is the high rate of discrimination between different original signals and their denoised versions. In the plots in Fig. 4, this fact is represented by the  $ds_{21}$ ,  $ds_{21\_denoised}$ ,  $ds_{12}$  and  $ds_{12\_denoised}$ . The distances of these measures are always higher than the distances between  $ds_{22\_denoised}$  and  $ds_{11\_denoised}$ . Even, the distances  $ds_{22}$  and  $ds_{11}$ , that is, the distances between one signal and its noisy versions are always lower than the cross distances, at least, for values of SNR higher than 2.



**Fig. 4.** Distance between signals (see Table 1 for legends)

## 5 Conclusions

Machine sound varies depending on factors such as background noise, failures of their mechanisms, environmental aspects (speech, superposition...), etc. Besides, when the feature sound is immersed in heavy perturbations as the previously cited is hard to identify. CWT can be used to discover the relevant signal components respect the selected wavelet bases. Then, using a proper basic wavelet, we can reconstruct the sound signal filtering the wavelet coefficients. The machine sound can be denoised following this procedure. Together with a specific feature extraction and identification (based on a basic distance measure) the noisy signals can be effectively identified, even with relatively low signal/noise ratio values.

## References

- [1] Mori, K., Kasashima, N., Yoshioha, T., Ueno, Y.: Prediction of Spalling on a Ball Bearing by Applying the Discrete Wavelet Transform to Vibration Signals. *Wear* 195(1-2), 162–168 (1996)
- [2] Roberts, S., Everson, R.: *Independent Component Analysis: Principles and Practice*. Cambridge Univ. Press, Cambridge, UK (2001)
- [3] Bolea, Y., Grau, A., Sanfeliu, A.: Non-speech Sound Feature Extraction based on Model Identification for Robot Navigation, 8 Iberoamerican Congress on Pattern Recognition. In: Sanfeliu, A., Ruiz-Shulcloper, J. (eds.) *CIARP 2003*. LNCS, vol. 2905, pp. 221–228. Springer, Heidelberg (2003)
- [4] Mallat, S., Zhang, Z.: Matching pursuits with time-frequency dictionaries. *IEEE Trans. on Signal Processing* 45(12), 3397–3415 (1993)
- [5] Donoho, D.-L.: De-noising by soft-thresholding. *IEEE Trans. on Information Theory* 33(7), 2183–2191 (1999)
- [6] Lin, J.: Feature Extraction of Machine Sound using Wavelet and its Application in Fault Diagnosis. *NTD&E International* 34, 25–30 (2001)
- [7] Goupilland, P., Grossmann, A., Morlet, J.: Cycle octave and related transforms in seismic signal analysis. *Geoexploration* 23, 85–102 (1984)
- [8] <http://spib.rice.edu/spib/data/signals/noise/babble>