Structural Pattern Recognition for Industrial Machine Sounds Based on Frequency Spectrum Analysis

Yolanda Bolea[1], Antoni Grau[1], Arthur Pelissier[1], and Alberto Sanfeliu[2]

[1] Automatic Control Dept, Technical University of Catalonia UPC, Barcelona, Spain
{yolanda.bolea,antoni.grau}@upc.es
[2] Institute of Robotics, IRI, UPC, Barcelona, Spain

Abstract. In order to discriminate different industrial machine sounds contaminated with perturbations (high noise, speech, etc.), a spectral analysis based on a structural pattern recognition technique is proposed. This approach consists of three steps: 1) to de-noise the machine sounds using the Morlet wavelet transform, 2) to calculate the frequency spectrums for these purified signals, and 3) to convert these spectrums into strings, and use an approximated string matching technique, finding a distance measure (the Levenshtein distance) to discriminate the sounds. This method has been tested in artificial signals as well as in real sounds from industrial machines.

1 Introduction

A common problem encountered in industrial environments is that the electric machine sounds are often contaminated by 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. For example, the effectiveness of the Fourier transform relies on the signals containing distinct characteristic frequency components of sufficient energy content, within a limited frequency band. If however the feature components spread over a wide spectrum, it can be difficult to differentiate them from other disturbing or masking components, especially when the feature components are weak in amplitude. This has been shown in various situations involving machine systems with incipient defects [1][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].

A. Sanfeliu et al. (Eds.): CIARP 2004, LNCS 3287, pp. 287–295, 2004.

In this work, we propose a new approach in order to discriminate among different industrial machine sounds, which can be affected by noise of various sources. We use the Morlet wavelet to de-noise the machine sounds, before frequency spectrums are extracted. These purified spectrums are the bases for a comparison between sound signals and a further discrimination step among sounds. A structural pattern recognition technique is used to compare the signal spectrums, because we convert each spectrum into a string, and a distance is found between strings. Since frequency spectrum does not follow a perfect pattern repeated along signals, it is not possible to use an exact matching algorithm to compare spectrums. To perform such comparison an approximated matching is used and the Levenshtein distance between spectrums is found. If the distance is short enough, these spectrums correspond to similar sounds. The use of string-to-string correction problem applied to pattern recognition is deeply treated in [7] and [8]. In order to check our approach, firstly we use some artificial signals with added gaussian noise, and the results are promising enough as to be used with real sounds.

This paper is organized as follows. In Section 2 the Morlet wavelet transform for de-noising the acoustical signals is explained. In Section 3 the approximated string matching is shown. Simulation and experimental results are presented in Section 4 and Section 5.

2 Wavelet and Its Application for Feature Extraction

2.1 Review of Wavelet Transform

The wavelet was originally introduced by Goupilland et al. in 1984 [9]. Let $\psi(t)$ be the basic wavelet function or the mother wavelet, then the corresponding family of daughter wavelets consists of

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

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

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

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

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

2.2 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 classical inversion formula is

$$x(t) = C_\psi^{-1} \int\int W(a,b)\psi_{a,b}(t)\frac{da}{a^2}db \tag{4}$$

Another simple inverse way is to use the Morlet's formula, which only requires a single integration. The formula is:

$$x(t) = C_\psi^{-1} \int W(a,b)\frac{da}{a^{3/2}} \tag{5}$$

where

$$C_{1\psi} = \int_{-\infty}^{\infty} \hat{\psi}^*(\omega)/|\omega| d\omega \tag{6}$$

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 purified by using formula Eq. (5). Thus, the key to obtaining the purified signal is how to obtain these feature coefficients.

Wavelet coefficients measure the similarity of the signal and each daughter wavelet. The more the daughter wavelet is similar to the feature component, the larger is the corresponding wavelet coefficient. So these large wavelet coefficients are mainly produced by the impulse components in the signal if the signal is transformed by the Morlet wavelet. We can get the impulse components in the signal reconstructing these large coefficients. Usually a threshold T_w should be set in advance, but it is not evident to choose it properly. 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.

3 Approximated Matching of Strings

Since we propose a structural approach, two purified frequency spectrum can be represented by R_p and R_q, the discrimination step is defined as follows: two sounds p and q are similar iif their purified frequency spectrum R_p and R_q approximate match.

The problem of string-matching can generally be classified into exact matching and approximate matching. For exact matching, a single string is matched against a set of strings and this is not the purpose of our work. For approximate string matching, given a string v of some set V of possible strings, we want to know if a string u approximately matches this string, where u belongs to a subset U of V. In our case, V is the global set of purified frequency spectrums and u and v are purified frequency spectrums obtained from different sounds. Approximate string matching is based on the string distances that are computed by using the editing operations: substitution, insertion and deletion [10].

Let Σ be a set of symbols and let Σ^* be the set of all finite strings over Σ. Let Λ denote the *null string*. For a string $A = a_1a_2...a_n \in \Sigma^*$, and for all $i, j \in \{1, 2,..., n\}$, let $A<i, j>$ denote the string $a_ia_{i+1}...a_j$, where, by convention $A<i, j> = \Lambda$ if $i > j$.

An *edit operation* s is an ordered pair $(a, b) \neq (\Lambda,\Lambda)$ of strings, each of length less than or equal to 1, denoted by $a \rightarrow b$. An edit operation $a \rightarrow b$ will be called an *insert* operation if $a = \Lambda$, a *delete* operation if $b = \Lambda$, and a substitution operation otherwise.

We say that a string B results from a string A by the edit operation $s = (a \rightarrow b)$, denoted by $A \rightarrow B$ via s, if there are strings C and D such that $A = CaD$ and $B= CbD$. An *edit sequence* $S:= s_1s_2...s_k$ is a sequence of edit operations. We say that S *takes* A *to* B if there are strings $A_0, A_1, ..., A_k$ such that $A_0 = A$, $A_k = B$ and $A_{i-1} \rightarrow A_i$ via s_i for all $i \in \{1, 2, ..., k\}$.

Now let γ be a cost function that assigns a nonnegative real number $\gamma(s)$ to each edit operation s. For an edit sequence S as above, we define the cost $\gamma(S)$ by $\gamma(S):= \Sigma_{i=1,...,k}\gamma(s_i)$. The *edit distance* $\delta(A, B)$ from string A to string B is now defined by $\delta(A, B):= \min\{\gamma(S) \mid S$ is an edit sequence taking A to $B\}$. We will assume that $\gamma(a \rightarrow b)= \delta(A, B)$ for all edit operations $a \rightarrow b$. The key operation for string matching is the computation of edit distance. Let A and B be strings, and $D(i,j)= \delta(A(1, i), B(1, j))$, $0 \leq i \leq m$, $0 \leq j \leq n$, where m and n are the lengths of A and B respectively, then:

$$D(i,j)= \min\{\, D(i\text{-}1,j\text{-}1) + \gamma(A(i) \rightarrow B(j)), D(i\text{-}1,j) + \gamma(A(i) \rightarrow \Lambda), D(i,j\text{-}1) + \gamma(\Lambda \rightarrow B(j)) \,\} \tag{7}$$

for all $1 \leq i \leq m$, $1 \leq j \leq n$. Determining $\delta(A, B)$ in this way can in fact be seen as determining a minimum weighted path in a weighted directed graph. Note that the arcs of the graph correspond to insertions, deletions and substitutions. The Levenshtein distance (metric) is the minimum-cost edit sequence taking A to B from vertices $v(0,0)$ to $v(n,m)$. In our case both strings have the same length (N) and the algorithm used is $O(N^2)$ [7].

4 Discrimination of Artificial Signals

In order to test the capacity of analysis, feature extraction, and discrimination of the above proposed method, eight artificial signals have been taken. The first set of signals are two sinusoidal signals with the equation $S_i(t)=A_i\cos(2\pi f_i t)$ (with i=1,2, and A_1=0.2, f_1=0.002Hz, and A_2=0.1, f_2=0.01Hz) and two signals described by the following expressions:

$$S_3(t) = 0.75\big(\exp(-(t-200)^2/2400)\cos(\pi t/6) + \exp(-(t-400)^2/3000)\cos(\pi t/5.4) + \exp(-(t-600)^2/2700)\cos(\pi t/7)\exp(-(t-800)^2/3200)\cos(\pi t/4.7)\big)$$

$$S_4(t) = \exp(-(t-200)^2/80)\cos(\pi t/15) + \exp(-(t-400)^2/70)\cos(\pi t/18) + \exp(-(t-600)^2/90)\cos(\pi t/14)\exp(-(t-800)^2/60)\cos(\pi t/16)$$

The other set of signals are the contaminated first set of signals, with additive white noise, which has a normal distribution with variance σ^2 = 0.2 and zero mean. The SNR for these signals is 0.09677, 0.02471, 0.36153 and 0.11263, respectively.

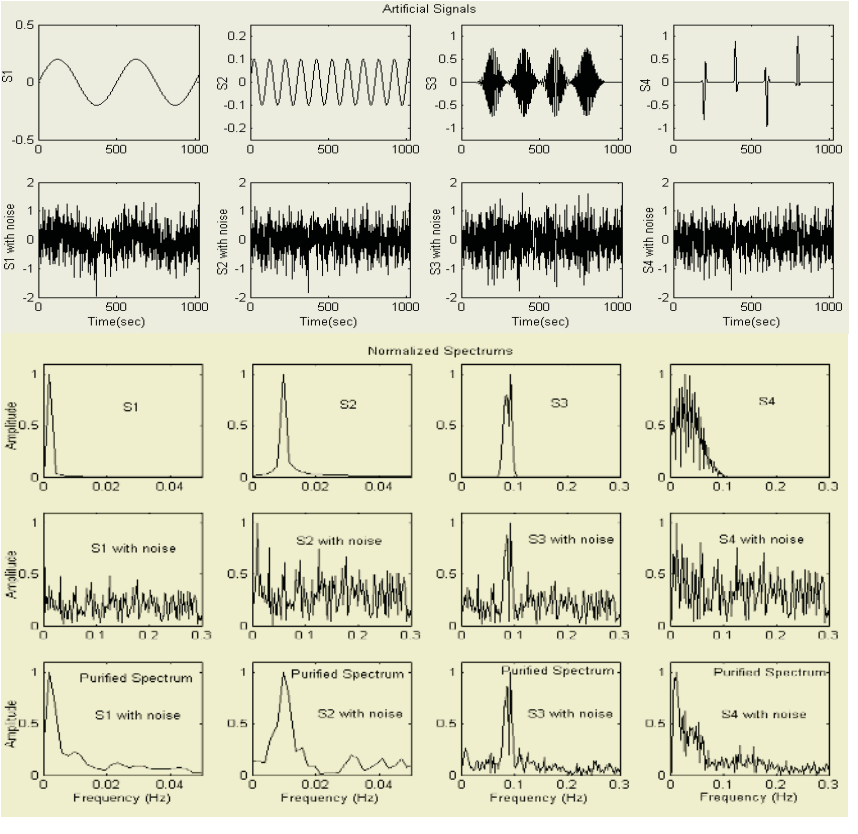

Fig. 1. Artificial signals; top to bottom: signals, signals with noise, frequency spectrum of the clean signals, frequency spectrum of contaminated signals, and frequency spectrum of the purified noisy signals.

Fig. 1 shows the two sets of artificial signals. In the first rows, the clean signals S_1, S_2, S_3 and S_4 as well as their contaminated versions can be seen. The next row contains the frequency spectrum from the clean signals, while the fourth row contains the spectrum of the signals with noise. It is important to note that there exists a huge difference between the signal spectrum with noise respect the spectrum of its clean signal.

In order to remove the maximum added noise to the signals, the contaminated signals are transformed with the Morlet wavelet, filtered (that is, removing the coefficients lower that a prefixed threshold) and reconstructed again. The frequency spectrum for these purified signals can be seen in the fifth row, and now, it is easy to observe that they are quite similar to these spectrums of the clean signals. The threshold T_w values are set to one-third of the maximum wavelet coefficients of the clean signals, fulfilling the basic rule stated in Section 2.2.

In order to quantify the similarity between signals, these spectrums are treated as strings, where each position is the amplitude of the spectrum. As the match will not be perfect, we use an approximate string matching technique and the Levenshtein distance (metric) is calculated, see Table 1. The distance is normalized respect the maximum distance, and the higher the distance, the more unlikely the signals is the same.

Table 1. Normalized distance from the artificial clean signals to the contaminated signals.

	Distance, using $T_s = 0$			
Clean signals	**Signals with noise**			
	S_1	S_2	S_3	S_4
S_1	0.32	0.60	0.68	0.92
S_2	0.39	0.56	0.68	0.88
S_3	0.54	0.69	0.40	1.00
S_4	0.58	0.71	0.81	0.64

The study and analysis of the discrimination algorithm are performed setting three threshold values T_s. This threshold T_s serves to eliminate all the amplitudes in the frequency spectrum above its value. Initially, we do not use any threshold ($T_s = 0$), and we use all the amplitudes in the spectrum; second, we use a value of $T_s = 0.2$ in order to remove the spurious frequencies; and finally, the threshold is set to $T_s = 0.5$ to capture the fundamental frequency and the most important harmonics in the signal.

In this study, we have realized that a similar distance i) between a clean signal and its contamined version and ii) between this clean signal and another contamined signal, can be discriminant enough if the distance between the clean signals is close. For this reason, it is important to check all the distances among the clean and contamined signals. This effect can be observed in Table 1. The distance between S_1 and S_1 with noise (d= 0.32) and S_2 and S_1 with noise (d=0.39) is enough to think that S_1 with noise is a contamined version of S_1, because the distance between S_1 and S_2 is only

d=0.09. The same reasoning is applied to S_2 with noise respect S_1 and S_2, obviously. In the other contamined signals (S_3 and S_4) their distance to their clean signals is short enough to perform a good discrimination (d=0.40 and d=0.64).

5 Discrimination of Real Sounds

For testing the proposed method with real sounds we have been working with 4 machine sounds: mill sound (S_{l1}), drill sound (S_{l2}), mill sound contamined with vibrations and speech (S_{r1}), drill sound contamined with speech (S_{r2}). The two former signals are considered the clean sounds. The latter are their contaminated version. The frequency sample is 22,050Hz, 16-bit, mono.

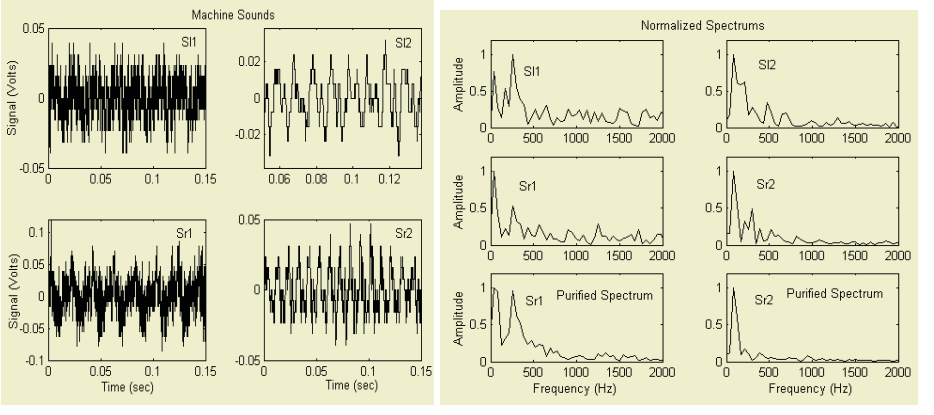

Fig. 2. (Left) Machine sounds; (right) first and second row: spectrums of machine sounds; third row: purified spectrums of the contaminated signals S_{r1} and S_{r2}.

As it can be seen in Fig.2. (1st and 2nd row right), the spectrum of clean signals (S_{l1} and S_{l2}) have two distinct features: i) S_{l1} has two important frequency peaks and S_{l2} has only one; ii) the fundamental frequencies of each signal are located at different spectrum positions. Taking into account these features and the low SNR of contaminated signals, the threshold T_w values are set to one-fourth of the maximum wavelet coefficient.

Table 2. Normalized distance among the real sounds and their contaminated versions.

	Distance	
Clean signals	**Signals with noise**	
	S_{r1}	S_{r2}
S_{l1}	0.58	0.64
S_{l2}	0.83	0.35

When the drill sound is contamined with vibrations and speech, the signal becomes stronger and there is a shift in the fundamental frequency from 258.7 Hz to 44.5 Hz due to: i) the vibrations reduce the fundamental frequency; and ii) the pitch in adult male speakers is between 50 Hz and 250 Hz. On the other hand, when the mill signal is contaminated with speech, the fundamental frequency does not vary because the clean signal fundamental frequency is already in this range (about 86.1 Hz).

In Fig.2 (3rd row right), the purified spectrum (with Morlet wavelet transform) captures very well the most important frequency peaks of the clean signals.

When these spectrums are converted into strings, if the frequency peaks between signals are closely located, the distance will also be close, and then the discrimination will be effective.

Many experiments have done with different parameters, $T_s = 0.2$ and $T_s = 0.6$, considering all the frequencies or discretizing them (the X-axis of the spectrums) when the string is generated. In all the cases the results (see Table 2) show that the proposed method can be used to discriminate real sounds.

6 Conclusions

Machine sound varies depending on the factors as background noise, failures of their mechanisms, environmental aspects (speech, noise, ...), etc. Besides, when the feature sound is immersed in heavy perturbations as the previously cited is hard to capture. CWT can be used to discover the relevant signal components respect the selected wavelet bases. Then, using a proper basic wavelet, we can obtain the feature components of a signal by reconstructing the wavelet coefficients. The machine sound can be purified following this procedure. Together with an approximated matching technique, the original source of real contamined sounds can be effectively detected.

References

1. Mori, K., Kasashima, N., Yoshioha, T. and Ueno, Y, “Prediction of Spalling on a Ball Bearing by Applying the Discrete Wavelet Transform to Vibration Signals”, Wear, vol.195, no.1-2, pp. 162-168, 1996.
2. Liu, H.-C. and Srinath, M.D., "Classification of partial shapes using string-to-string matching", *Intell. Robots and Comput. Vision*, *SPIE Proc.*, vol. 1002, pp. 92-98, 1989.
3. Bolea, Y., Grau, A. and Sanfeliu, A., “Non-speech Sound Feature Extraction based on Model Identification for Robot Navigation”, *8th Iberoamerican Congress on Pattern Recognition*, CIARP 2003, Lectures Notes in Computer Science, LNCS 2905, pp. 221-228, Havana, Cuba, November 2003.
4. Mallat, S. and Zhang, Z., “Matching pursuits with time-frequency dictionaries”, *IEEE Trans. on Signal Processing*, vol.45, no.12, pp. 3397-3415, 1993.
5. Donoho, D.-L., “De-noising by soft-thresholding”, *IEEE Trans. on Information Theory*, vol.33, no.7, pp. 2183-2191, 1999.
6. Lin, J., “Feature Extraction of Machine Sound using Wavelet and its Application in Fault Diagnosis”, NTD&E International, vol.34, pp.25-30, 2001.

7. Sankoff, D. and Kruskal, J.B. eds, *Time Warps, String Edit and Macromolecules: The Theory and Practice of Sequence Comparison*, Addison-Wesley, Reading, MA, 1983.
8. Bunke, H. and Sanfeliu, A., *Syntactic and Structural Pattern Recognition Theory and Applications*, Series in Computer Science, vol.7, World Scientific Publ., 1990.
9. Goupilland, P., Grossmann, A. and Morlet, J., “Cycle octave and related transforms in seismic signal analysis”, *Geoexploration*, vol.23, pp.85-102, 1984.
10. Wagner, R.A. et al., "The string-to-string correction problem", *J. Ass. Comput. Mach.*, vol.21, no.1, pp. 168-173, 1974.