Higher-order statistics to detect and characterise termite emissions

J.J.G. de la Rosa, I. Lloret, C.G. Puntonet and J.M. Górriz

An independent components analysis algorithm is applied to extract vibratory signals generated by termites from background noise. Signals from a microphone in uniform and Gaussian noise, were taken as sensor outputs. Bispectrum is proposed as a higher-order statistic to characterise time series. Experiments prove that detection and characterisation can be performed successfully even with low signal-to-noise ratio signals.

Introduction: The costs of the harm caused by termites could be reduced through earlier detection. Detection is also important because environmental laws are becoming more restrictive with termiticides. Besides, only about 25% of the affected structure is accessible [1]. New techniques have been developed to gain accessibility.

Acoustic signals produced when wood fibres are broken by termites can be monitored using acoustic emission (AE) sensors, targeting infestations by means of spectral and temporal analysis. Their drawback is the relative high cost and biophysical limitations.

The aim in this Letter consists of using higher-order statistics (HOS) for a twofold purpose. First, an independent components analysis (ICA) cumulant-based algorithm is used to separate alarm signals from additive stationary noise. This could be the basis of separating low-level termite signals from urban noise using cheap equipment. Secondly, the bispectrum has been applied to obtain an improved characterisation of emissions in the frequency domain.

A previous estimation of the power spectrum of termite emissions was developed using a seismic accelerometer, with the aim of getting a biological reference. Data were acquired in Málaga (Spain), in subterranean wood structures and roots.

Acoustic detection of termites: characteristics and devices: When disturbed in their galleries, soldiers produce vibratory signals by drumming their heads against the substratum [1]. The signals consist of pulse trains with pulse repetition rates in the range of 10–25 Hz, with burst rates around 500–1000 ms, depending on the species [1]. The amplitudes of signals are highly variable and depend on the wood and strength of the taps. Although these vibratory signals have distinctive time instances it is difficult to detect them in a noisy environment. A variety of signal processing methods have been used in similar situations in other fields of science and technology [1–6]. They include statistical, spectral and time-frequency analysis combined with wavelets. They are all based on energy conservation, being only useful for finding predominant information. As a consequence, low signal-to-noise ratio sources cannot be identified successfully. Spurious events of interest, like pulse-like events, are buried. HOS and ICA bring a different strategy in dealing with source separation and identification of non-Gaussian random processes.

ICA model, cumulants and polyspectra: Blind source separation (BSS) by ICA is receiving attention because of its numerous applications in many fields such as speech recognition and medicine [2, 6]. Let $s(t) = [s_1(t), s_2(t), \ldots, s_n(t)]'$ be the vector of unknown sources, where the superscript represents *transpose*. The known mixtures are modelled by

$$x(t) = As(t) \tag{1}$$

where $x(t) = [x_1(t), x_2(t), \ldots, x_m(t)]'$ is the vector of observations and $A = [a_{i,j}] \in \Re^{m \times n}$ is the unknown mixing matrix, modelling the environment in which signals are mixed, transmitted and measured. We assume that A is a non-singular $n \times n$ square matrix. The goal of ICA is to find an $n \times m$ separating matrix B such that it extracts source signals via

$$\hat{s} = y(t) = Bx(t) = BAs(t) \tag{2}$$

where $y(t) = [y_1(t), y_2(t), \ldots, y_n(t)]'$ is the separated source vector which is an estimator of the original vector of sources [2–4]. If the complete determination of the mixing matrix A were possible, BA would be the identity.

Cumulants and polyspectra reveal information about amplitude and phase, whereas second-order statistics are phase-blind [2–6]. For example, the fourth-order cumulant of a set of variables is given by:

$$Cum(x_1, x_2, x_3, x_4) = E\{x_1x_2x_3x_4\} - E\{x_1x_2\}E\{x_3x_4\} - E\{x_1x_3\} \times E\{x_2x_4\} - E\{x_1x_4\}E\{x_2x_3\} \tag{3}$$

Let $\{x(t)\}$ be a rth-order stationary random process. The rth-order cumulant is defined as the joint rth-order cumulant of the random variables $x(t), x(t+\tau_1), \ldots, x(t+\tau_{r-1})$:

$$C_{r,x}(T_1, T_2, \ldots, T_{r-1}) = Cum[x(t), x(t+T_1), \ldots, x(t+T_{r-1})] \tag{4}$$

The higher-order spectra are usually defined in terms of the rth-order cumulants as their $(r-1)$-dimensional Fourier transforms

$$S_{r,x}(f_1, f_2, \ldots, f_{r-1}) = \sum_{\tau_1=-\infty}^{\tau_1=+\infty} \cdots \sum_{\tau_{r-1}=-\infty}^{\tau_{r-1}=+\infty} C_{r,x}(T_1, T_2, \ldots, T_{r-1}) \times \exp[-j2\pi(f_1T_1 + f_2T_2 + \cdots + f_{r-1}T_{r-1})] \tag{5}$$

The special polyspectra derived from (5) are power spectrum ($r=2$), bispectrum ($r=3$) and trispectrum ($r=4$). To extract useful information one-dimensional slices of cumulant sequences and polyspectra are used in non-Gaussian stationary processes.

ICA algorithm: Separation of the sources can be developed using a contrast function based on the entropy of the outputs [2]. Using this function it can be shown [2, 4] that the separating matrix can be obtained by means of the recurrent equation

$$B^{(h+1)} = \lfloor I + \mu^{(h)}(C^{1,\beta}_{y,y}S^{\beta}_{y} - I)\rfloor B^{(h)} \tag{6}$$

where S^{β}_{y} is the matrix of the signs of the output cumulants. Equation (7) can be interpreted as a quasi-Newton algorithm of the cumulant matrix $C^{1,\beta}_{y,y}$. The learning rate $\mu^{(h)}$ is described in [4]. Convergence of (6) is reached if the matrix $C^{1,\beta}_{y,y}S^{\beta}_{y}$ tends to the identity.

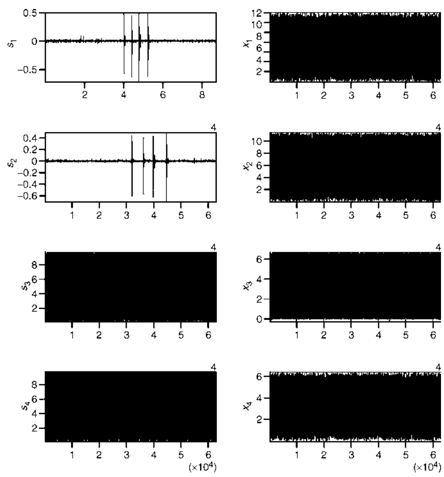

Fig. 1 *Sources and their mixtures*

Horizontal units 1/96000, s

Results: A low-cost microphone, Ariston CME6 model, with a sensitivity of 62 dB and a bandwidth of 100 Hz to 8 kHz was connected to the sound card of a portable computer. Sources 1 and 2 consist of two zero-mean normalised bursts. The third and fourth sources consist of two uniform distributed noise signals with enough amplitude to mask the burst. The elements of the 4×4 mixing matrix are chosen from uniformly distributed random numbers between 0 and 1. A comparison between the impulse response of the accelerometer (KB12V, MMF) and the spectrum of one impulse in a burst was performed. Significant drumming responses are produced over the range 200 Hz to 4 kHz and the carrier frequency is around 2600 Hz [1, 2]. The spectrum is not flat against frequency as one would expect for a pulse-like event. Fig. 1 shows the original sources and the mixtures, which give very little information about the original signals. With the sources, a number of differences are found. First, the carrier in the spectra of the separated signals, $y_1(t)$ and $y_2(t)$, matches the carrier frequency in the spectra of the impulsive response of the accelerometer.

To provide some insight about the power of the higher-order statistics for characterisation purposes, ten replications of signal-plus-white Gaussian noise were generated (SNR = 0 dB). The average diagonal bispectrum is shown in Fig. 2. We observe very sharp peaks of the energy, concentrated in a narrow range of frequencies. These one-diagonal measures underlie information concerning the phase coupling of harmonics at integer multiples of the fundamental one.

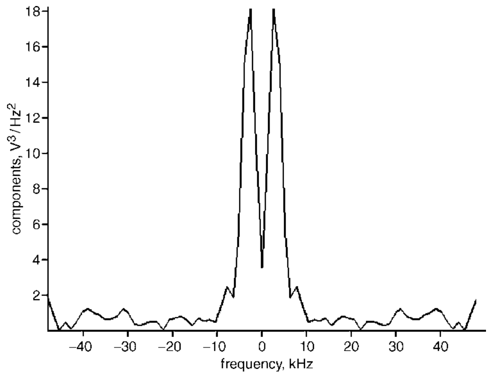

Fig. 2 *Diagonal average bispectrum of masked alarm signals*

Conclusions: This work shows that the ICA algorithm separates sources with small energy levels in comparison to the background noise. This is explained by the statistical independence basis, regardless of the energy associated to each frequency component. Results of the spectra lead us conclude that the separation has been performed correctly, because the same spectral shape as the accelerometer response is outlined. In this stage we have proved the validity of ICA over a pre-processed set of signals. A time-domain characterisation is enough. We have also presented a HOS-based method of characterisation of the vibratory signals. Application of the polyspectra of the bispectra has enhanced the characterisation of the data sequences in Gaussian noise.

Acknowledgments: We acknowledge the Spanish Ministry of Science and Technology for funding the project DP12003-00878, and the Andalusian Government for funding part of this research.

30 June 2004

Electronics Letters online no: 20045664

doi: 10.1049/el:20045664

J.J.G. de la Rosa and J.M. Górriz (*Electronics Instrumentation Research Group, Engineering School of Algeciras, University of Cádiz Avda, Ramón Puyol S/N, Algeciras, Cádiz 11202, Spain*)

I. Lloret (*Department of Computer Science, Engineering School of Algeciras, University of Cádiz Avda, Ramón Puyol S/N, Algeciras, Cádiz 11202, Spain*)

C.G. Puntonet (*Department of Architecture and Computers Technology, University of Granada, ESII C/Periodista Daniel Saucedo, Granada 18071, Spain*)

References

1 Robbins, W.P., Mueller, R.K., Schaal, T., and Ebeling, T.: 'Characteristics of acoustic emission signals generated by termite activity in wood'. IEEE Ultrasonic Symp., Orlando, FL, USA, 1991, pp. 1047–1051

2 Puntonet, C.G., and Mansour, A.: 'Blind separation of sources using density estimation and simulated annealing', *IEICE Trans. Fundam. Electron. Commun. Comput. Sci.*, 2001, **E84-A**

3 Chambers, J.A., Jafari, M.J., and McLaughlin, S.: 'Variable step-size EASI algorithm for sequential blind source separation', *Electron. Lett.*, 2004, **40**, (6), pp. 393–394

4 Górriz, J.M., Puntonet, C.G., de la Rosa, J.J.G., and Salmerón, M.: 'Time-series forecasting using RBFs and exogenous data, intelligent systems design and applications', ser. Advances in Soft-Computing (Springer-Verlag, Berlin/Heidelberg/New York), 2003, p. XVIII

5 Jakubowsky, J., Kwiatos, K., Chwaleba, A., and Osowsky, S.: 'Higher order statistics and neural network for tremor recognition', *IEEE Trans. Biomed. Eng.*, 2002, **49**, (2), pp. 152–159

6 Mendel, J.M.: 'Tutorial on higher-order statistics (spectra) in signal processing and system theory: theoretical results and some applications', *Proc. IEEE*, 1991, **79**, (3), pp. 278–305