Modern Physics Letters B, Vol. 25, Nos. 12 & 13 (2011) 1133–1142

DOI: 10.1142/S021798491102684X

PERMUTATION ENTROPY APPLIED TO MOVEMENT BEHAVIORS OF *DROSOPHILA MELANOGASTER*

YUEDAN LIU and TAE-SOO CHON*

Department of Biological Sciences, Pusan National University,
Busan (Pusan) 609-735, Korea
**tschon@pusan.ac.kr*

HUNKI BAEK

Department of Mathematics Education, Catholic University of Daegu,
Gyeongsan, Gyeongbuk 712-702, Korea

YOUNGHAE DO

Department of Mathematics, Kyungpook National University, Daegu 702-701, Korea

JIN HEE CHOI

Faculty of Environmental Engineering, College of Urban Science,
University of Seoul, Seoul 130-743, Korea

YUN DOO CHUNG

Department of Life Science, College of Natural Science,
University of Seoul, Seoul 130-743, Korea

Movement of different strains in *Drosophila melanogaster* was continuously observed by using computer interfacing techniques and was analyzed by permutation entropy (PE) after exposure to toxic chemicals, toluene (0.1 mg/m^3) and formaldehyde (0.01 mg/m^3). The PE values based on one-dimensional time series position (vertical) data were variable according to internal constraint (i.e. strains) and accordingly increased in response to external constraint (i.e. chemicals) by reflecting diversity in movement patterns from both normal and intoxicated states. Cross-correlation function revealed temporal associations between the PE values and between the component movement patterns in different chemicals and strains through the period of intoxication. The entropy based on the order of position data could be a useful means for complexity measure in behavioral changes and for monitoring the impact of stressors in environment.

Keywords: Movement patterns; cross-correlation; behavioral monitoring; strains; toluene; formaldehyde.

*Corresponding author.

1. Introduction

The objective and automatic computational detection of response behaviors of indicator species has been considered as an efficient means of *in situ* biomonitoring.[1] Behavioral monitoring is essential in filling the gap between macro-scale (e.g. community structure) and micro-scale (e.g. molecular analysis) measurements in order to establish integrative assessment systems across different time and space scales.[2] Detection of response behaviors has additional advantages such as minimization of environmental damage, economy and efficiency in measurement, and a possibility of continuous *in situ* monitoring. Behavioral monitoring has been reported on various indicator taxa such as fishes,[3,4] crustaceans[5] and insects[6] by utilizing computational methods including correlation analysis,[7] fractal dimension8 and hidden Markov model.[9–11]

In coping with elucidation of the time-space structured property in behavioral data, we applied a complexity measure, permutation entropy (PE),[12] to detect behavioral changes in response to stressors. PE was introduced as a fast and robust algorithm to investigate time series data in response to stimuli mainly in medicine, including fetal heart rate,[13] electroencephalography,[14–16] and vigilance of patients.[17] However, PE has never been applied to response behaviors of indicator species treated with toxic chemicals. In this study we focused on effect of the order of position data in elucidating internal (i.e. strains) and external (i.e. chemicals) constraints on time series response behaviors. The continuous positional data were used for calculating PE to reflect changes in movement behaviors in response to constraints. Temporal associations between PE and component movement patterns were further checked through cross-correlation function to reveal the time delayed effects of chemicals in the transient periods after exposure to stressors.

2. Materials and Method

2.1. *Test organisms and observation system*

Drosophila melanogaster was used as a test species since it has been considered as an indicator species regarding availability of biological information and vulnerability to chemical stressors.[18] Three strains, one wild type (s7) and two mutant strains (p38b, p53), were obtained from the Department of Life Science, College of Natural Science, University of Seoul, Korea, and were maintained of a temperature of $25 \pm 1°C$ and humidity $70\% \pm 5\%$ (average standard deviation) according to the standard rearing methods.[19] Test organisms were randomly chosen from stock populations and were placed individually in an observation cage (150 mm $\times$ 50 mm $\times$ 10 mm) in vertical position.

Two volatile organic compounds, toluene and formaldehyde (Sigma-Aldrich®), were treated to the test organisms in this study. The chemicals were diluted and sprayed on the cotton placed within the observation cage at the concentration of 0.1 mg/m^3 and 0.01 mg/m^3 respectively. Fifteen organisms for each group (3 strains

$\times$ 3 treatments) were observed separately for 6 hours without treatment and for 6 hours after the treatments of different chemicals.

The movement tracks of test organisms were recorded in two dimensions by using an observation system consisting of an observation cage, a CCD camera (Hitachi KP-D 20 BU®), a timer, an A/D interface (Matrox Morphis®), and an image recognition system (0.25 s/frame).[4,10,11]

2.2. *Permutation entropy (PE)*

PE is a measure of complexity for time series operating the ranks of the neighboring data.[12] Time series position data were used to present behavioral states of indicator organisms. Considering that vertical position is meaningful in presenting response behaviors of *D. melanogaster*,[18] one-dimensional vertical position data were used in this study: $\{y_t\}_{t=1,\dots,T}$, where y_t is the position (mm) in vertical direction at time t in the total sampling time of T. The data were classified as π according to the order of position with the number of m. Through preliminary experiments three consecutive positions with 0.25 s/frame ($m = 3$) were feasible to present the movement trends of *D. melanogaster* and were in accord with other reports related to neural and behavioral responses.[10,16] Suppose the time series data of vertical location of the observed individual are $y_t = (2, 5, 8, 12, 3, 6, 7, 11, 4, 10)$ with $T = 10$ and $m = 3$. The eight sets consisting three consecutive values, including (2, 5, 8), (5, 8, 12), (3, 6, 7) and (6, 7, 11), would be represented by the same increasing order and were classified as one type in π with the relative order of 012. Similarly, the position data, (8, 12, 3) and (7, 11, 4), could correspond to another type with the order of 120, regarding $y_{t+2} < y_t < y_{t+1}$. The permutations in different orders such as 012 or 120 were defined as the component movement patterns of *D. melanogaster* in this study. For each π, the relative frequency could be determined as (# meaning the number):[12]

$$p(\pi) = \frac{\#\{t | t \le T - m, (y_{t+1}, \dots, y_{t+m}) \text{ has type } \pi\}}{T - m + 1}\,. \tag{1}$$

In case equal values were observed in the time series position data, random numbers with accuracy smaller than measurements were given to the data. The order of the data of equal values was subsequently determined according to the given values of the random number.[13] The permutation entropy is consequently defined as

$$H(m) = -\sum p(\pi) \log p(\pi)\,, \tag{2}$$

where the sum runs over all permutations of order m. The PE values were measured from 15 organisms and the averages in 1 minute intervals were obtained for 6 hours "without" and "after" the treatments of chemicals in each strain. For convenience $H(m)$ was normalized by its maximum value $\log m!$:

$$h_m = H(m) / \log m!\,. \tag{3}$$

The normalized permutation entropy for the example stated above could be calculated as $h(3) = -((4/8)\log(4/8) + 2(2/8)\log(2/8))/\log 3! = 0.58$.

Cross-correlation function[20] was further obtained in revealing temporal associations between PEs and between proportions of component movement patterns during the transient period after the treatments. The cross-correlation function was normalized according to coefficient estimation in which the autocorrelations are 1.0 at zero lag. Calculation of PE and cross-correlation function was programmed and conducted under Matlab environments (The Mathworks, R2010b).

3. Results

3.1. *General movements*

The movement tracks were variable according to the different strains without the treatments (Fig. 1). Comparing with the tracks of the wild type s7 (Fig. 1(a)) and the strain p53 (Fig. 1(g)), movement of the strain p38b (Fig. 1(d)) was most active, spanning a wide range of the arena. Speed was different according to the strains without the treatments and consistently decreased in all strains after the treatments (Fig. 2(a)). The impact was stronger in p38b and with the treatment of formaldehyde (Fig. 1(f)). Two-factor analysis of variance showed the significance of speed between strains ($DF = 2$, $P < 0.001$ with $F = 965.8$), between chemicals ($DF = 2$, $P < 0.001$ with $F = 21352$), and interaction of strains and chemicals ($DF = 4$, $P < 0.001$ with $F = 549.0$).[21]

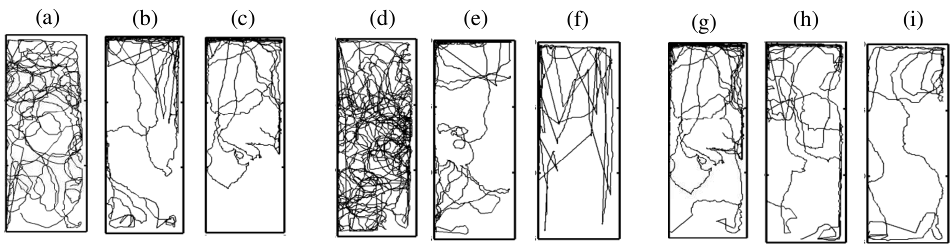

Fig. 1. The movement tracks of *D. melanogaster* (10 minutes). s7: (a) without treatment, and (b) after treatment with toluene and (c) formaldehyde, p38b: (d) without treatment and (e) after treatment with toluene and (f) formaldehyde, p53: (g) without treatment, and (h) after treatment with toluene and (i) formaldehyde.

3.2. *Permutation entropy and movement patterns*

The PE values accordingly responded to different strains and chemicals (Fig. 2(b)). In comparison with the wild type, s7, PE was higher in the strain p53 and lower in the strain p38b without the treatments. In contrast to the change in speed (Fig. 2(a)), however, PE values increased in all strains after the treatments, especially treated with formaldehyde (Fig. 2(b)).

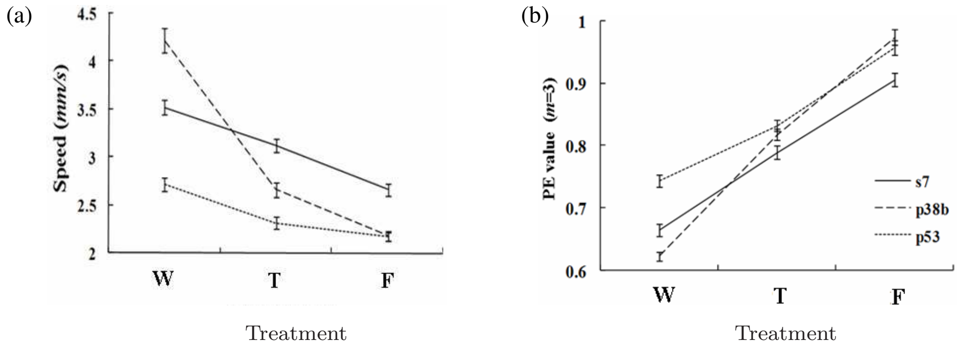

Fig. 2. Mean and standard deviation of speed and PE in different strains without and after the treatments of chemicals ($n = 15$). (a) Speed and (b) PE (W: without treatment; T: toluene; F: formaldehyde).

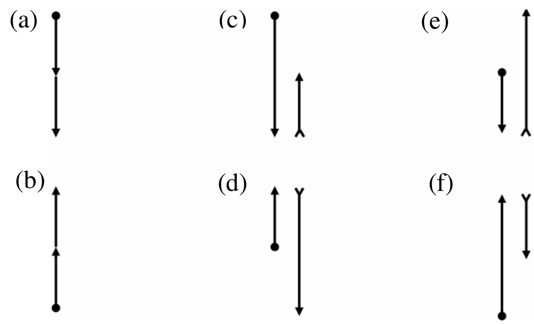

Fig. 3. Movement patterns of *D. melanogaster* based on ordering of consecutive vertical positions with $m = 3$ (0.50s). (a) down-down (210, P1); (b) up-up (012, P2); (c) down-up (201, P3); (d) up-down (120, P4); (e) down-up (102, P5) and (f) up-down (021, P6). (The length of the bar indicates relative distance in each movement. The dot in the bar stands for the starting point.)

According to the order of vertical position, six different movement patterns can be defined (Fig. 3). More irregular movements including the reverse movements (e.g. up-down movement; P4 and P6) appeared in a higher frequency after the treatments (Fig. 4). The continuous forward or backward movements that were strongly dominant without the treatments decreased substantially after the treatments. Although the movement tracks appeared to be longer (i.e. due to high speed) and more complex without the treatments in all strains (Figs. 1(a), 1(d) and 1(g)), the movement segments were in fact strongly dominated by the simple directional movement patterns (e.g. continuous forward movement) at this stage. The frequency of the various reverse movements (P3–P6 in Fig. 4) increased after the treatments. Consequently PE, presenting diversity in movement patterns (Eq. (2)), was higher after the treatments (Fig. 2(b)). Among the three strains, the increase in PE was most outstanding in the strain p38b. Formaldehyde was stronger in affecting PE values than toluene. Two-factor analysis of variance showed the significance of PE

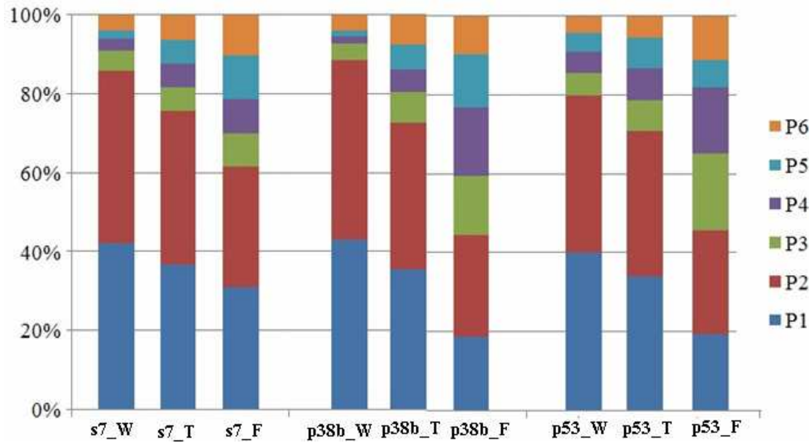

Fig. 4. Percentage of movement patterns of *D. melanogaster* defined by PE without and after the treatments ($m = 3$). (Pattern names are listed in Fig. 3.)

between strains ($DF = 2$, $P < 0.001$ with $F = 768.8$), between chemicals ($DF = 2$, $P < 0.001$ with $F = 13970$), and interaction of strains and chemicals ($DF = 4$, $P < 0.001$ with $F = 351.6$).[21]

The proportions of six component movement patterns accordingly varied in different strains after the treatments (Fig. 4). The movement patterns, P1 and P2, for instance, were most dominant in p38b without the treatments, while the proportions of these straight movement patterns were lower than the proportions observed in the strains s7 and p53 after the treatments. Chi-square test of three-dimensional contingency tables (strains $\times$ chemicals $\times$ movement patterns) showed the component movement pattern was not independent of chemicals and strains ($v = 40$, $P < 0.001$ with $\chi^2 = 424.5$). When the effects of chemicals on the movement patterns were tested separately in each strain, the proportions of the movement patterns were all dependent on different chemicals according to Chi-square test ($DF = 5$, $P < 0.001$ with $\chi^2 = 96.5$ for s7, $\chi^2 = 158.2$ for p38b, and $\chi^2 = 63.5$ for p53).[21] Consequently, the PE values and the proportions of component movement patterns were feasible in differentiating external (i.e. chemicals) and internal (i.e. strains) constraints according to complexity residing in movement data.

3.3. *Cross-correlation*

Cross-correlation of PE was analyzed in order to reveal the time delayed effects between different chemicals through the period of intoxication (Figs. 5(a)–5(b)).[20] The time series PE values of "without" the treatments (W) were cross-correlated with the PE values "with" the treatments of formaldehyde (F) and toluene (T) across the time lag (-6 h $\sim$ $+6$ h) (Fig. 5(a)). It was notable that cross-correlation

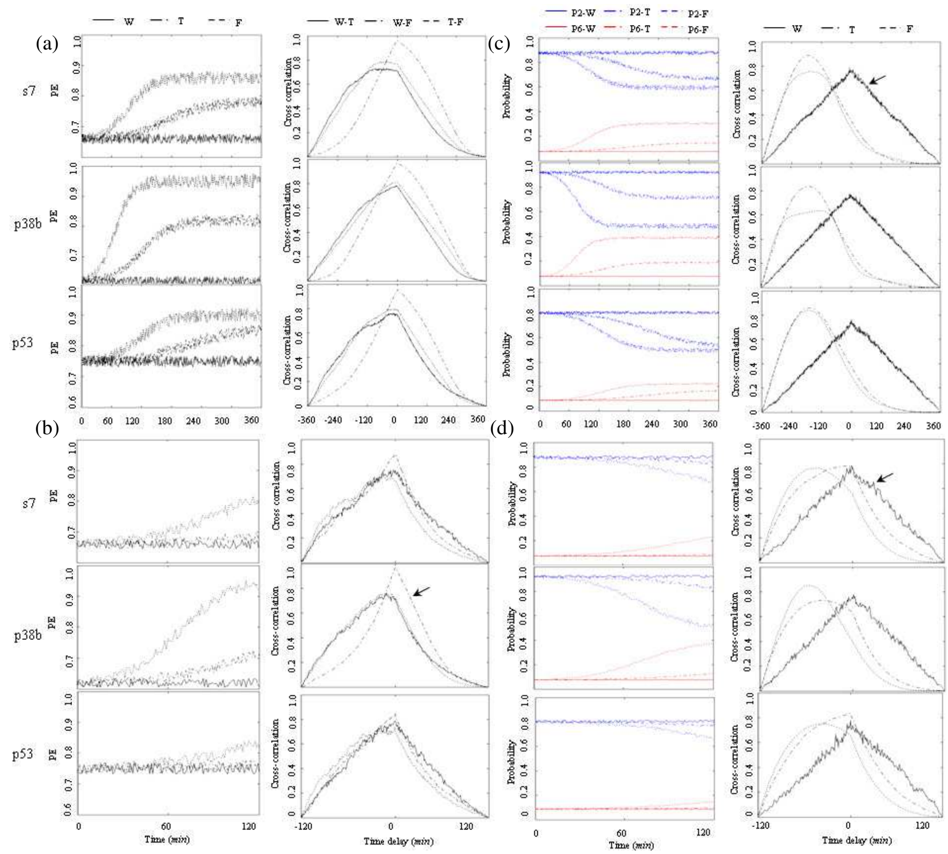

Fig. 5. Cross-correlation across different time delay in different strains and chemicals: (a) permutation entropy for $0 \sim 6$ h; (b) permutation entropy for $0 \sim 2$ h; (c) frequency of movement patterns P2 and P6 for $0 \sim 6$ h; (d) frequency of patterns P2 and P6 for $0 \sim 2$ h. (Subfigures in the left; time development of PE or frequency of movement patterns, and subfigures in the right; cross-correlation function with time delay. See text for explanation of the arrows.)

between W and F (W–F) addressed the slightly positive time lag while temporal associations between W and T (W–T) were minimal (i.e. linear decrease in cross-correlation function along with increase in time lag from the zero time delay) (Fig. 5(a)). Temporal association between two chemicals, T–F, was also minimal. This trend of cross-correlation was commonly observed for all the strains (Fig. 5(a)). When the data were checked in the early period ($0 \sim 2$h), temporal associations mostly disappeared except the slightly positive time delay in W–F in the strain of p38b (arrow in Fig. 5(b)).

Temporal associations were also observed in the frequency (number of occurrence per minute) of two component movement patterns, P2 and P6 (Figs. 5(c) and 5(d)). The patterns, P2 and P6, were representative movements for normal and intoxicated states, respectively (Fig. 4). The time series frequency data of P2 were cross-correlated with the data of P6 for "without" the treatments (W), and "with" the chemical treatments (T and F) across the time lags (-6 h $\sim$ $+6$ h and -2 h $\sim$ $+2$ h) (Figs. 5(c) and 5(d)).

The negative time delay was observed consistently with the chemical treatments for toluene (T) and formaldehyde (F) during the whole (Fig. 5(c)) and early (Fig. 5(d)) periods. Temporal association between P2 and P6 without the treatments (W), however, was minimal (i.e. decrease in cross-correlation function away form the zero time-delay) (arrows in Figs. 5(c) and 5(d)). This indicated existence of temporal associations between the two movement patterns (P2 and P6) with the chemical treatments.

4. Discussion and Conclusion

The PE values and frequencies of the component movement patterns accordingly presented effects of internal (i.e. strains) and external (i.e. chemicals) constraints (Figs. 2–4). It was understandable that PE increased after the chemical treatments (Fig. 2(b)), while speed decreased after the treatments (Fig. 2(a)). PE was able to illustrate diversity residing in the movement patterns covering both healthy and intoxicated states in response to chemicals at low concentrations (Figs. 2(b) and 4). Consequently, PE based on ordering of position expanded the scope of quantifying behavioral states in addition to conventional parameters based on distances (e.g. speed).

PE and frequency of the component movement patterns were useful for detecting changes in behavioral states in both overall and specific changes including the transient periods under stressful conditions (Figs. 2, 4 and 5). Conventional parameters such as speed and fractal dimension[8] were used for behavioral monitoring but mainly presented the overall trend of behavioral changes, being limited in presenting both global and local information pertaining to movement changes after exposure to stimuli.

It was notable that temporal associations between the component movement patterns (P2 and P6) were obtained with the negative time delay (Figs. 5(c) and 5(d)). Considering that the time delay effects were observed both in the early and whole periods, temporal associations between the components patterns seemed to be more sensitive in presenting behavioral changes compared with the case of the PE values (Figs. 5(a) and 5(b)). The high values in cross-correlation may indicate feasibility in monitoring response behaviors exposed to stressors in the earlier period (Fig. 5(d)). The consistency in temporal association appearing across different strains and chemicals was also notable. This may reflect an expression of a fixed network in physiology in presenting the responding behaviors after intoxication. Further study may be needed in revealing toxicological responses in the physiology-behavior relationships in the future. Cross-correlation in PE was additionally sensitive in the strain p38b in the early period (Fig. 5(b)). This type of the strain-specific response would be also meaningful in illustrating effects of the genes on behaviors. The results presented in this study could serve as a step stone for development of genomics of indicator species specifically responding to the selected stressors.

Considering that location information (i.e. order of position) is embedded in PE, the PE values are natural in presenting the "shape" information in movement behaviors. The complexity residing in movements was accordingly elucidated in differentiating effects of internal and external stimuli even in 1D data (Figs. 2–5). It would be additionally useful in quantifying the movement data if higher dimension is used for presenting more complex behaviors. In this study, we only used 1D data as an initial step of study, PE would be more suitable in presenting diverse behavioral changes if 2D or 3D data are used in the future.

In conclusion, permutation entropy was suitable in presenting behavioral states in different strains and chemicals. PE values and frequency of component movement patterns could be useful for monitoring response behaviors of indicator species, expanding the scope of computation in behavioral monitoring. Cross-correlation function applied to PE and the component movement patterns would be also feasible in illustrating temporal association of behavioral states and could be an efficient means of detecting time development of response behaviors in the transient periods after exposure to stressors.

Acknowledgment

This subject is supported by Ministry of Environment as "The Eco-technopia 21 project, Korea".

References

1. D. Lemly and R. J. Smith, *Ecotoxicol. Environ. Saf.* **11** (1986) 210.
2. A. Gerhardt, *Biomonitoring of Polluted Water — Reviews on Actual Topics*, ed. Gerhardt (Res. Forum, TTP, Zurich, 1999), Vol. 9, p. 95.
3. H. Dutta, J. Marcelino and C. Richmonds, *Arch. Intern. Physiol. Biochem. Biophys.* **100** (1992) 331.
4. T.-S. Chon, N.-I. Chung, I.-S. Kwak, J.-S. Kim, S.-C. Koh, S.-K. Lee, J.-B. Leem and E. Y. Cha, *Environ. Monit. Assess.* **101** (2005) 1.
5. S. D. Roast, J. Widdows and M. B. Jones, *Aquat. Toxicol.* **47** (2000) 227.
6. A. Gerhardt, A., l. Janssens de Bistoven and A. M. V. M. Soares, *Environ. Contam. Toxicol.* **48** (2005) 450.
7. W. Alt, in *Biological Motion*, eds. W. Alt and G. Hoffmann (Springer-Verlag, Berlin, 1990), p. 254.
8. C. W. Ji, S.-H. Lee, K.-H. Choi, I.-S. Kwak, S. G. Lee, E. Y. Cha, S.-K. Lee and T.-S. Chon, *Int. J. Ecodyn.* **12** (2007) 1.
9. A. Franke, T. Caelli and R. J. Hudson, *Ecol. Model.* **173** (2004) 259.
10. Y. Liu, T.-S. Chon and S.-H. Lee, *J. Korean Phys. Soc.* **56** (2010) 1003.
11. Y. Liu, S.-H. Lee and T.-S. Chon, *Ecol. Model.* (2011), doi:10.1016/j.ecomodel.2011.02.010.
12. C. Bandt and B. Pompe, *Phys. Rev. Lett.* **88** (2002) 174102.

13. B. Frank, B. Pompe, U. Schneider and D. Hoyer, *Med. Biol. Eng. Comput.* **44** (2006) 179–187.
14. X. Li, S. Cui and L. J. Voss, *Anesthesiology* **109** (2008) 448.
15. Y. Cao, W. Tung, J. B. Gao, V. A. Protopopescu and L.M. Hively, *Phys. Rev. E* **70** (2004) 46217.
16. E. Olofsen, J. W. Sleigh and A. Dahan, *Brit. J. Anaesth.* **101** (2008) 810.
17. A. A. Bruzzo, B. Gesierich, M. Santi, C. A. Tassinari, N. Birbaumer and G. Rubboli, *Neurol. Sci.* **29** (2008) 3.
18. K. M. Beckingham, J. D. Armstrong, M. J. Texada1, R. Munjaal and D. A. Baker, *Gravitational Space Biol.* **18** (2005) 17.
19. M. Ashburner and J. N. Thompson, in *The Genetics and Biology of Drosophila*, eds. M. Ashburner and T. R. F. Wright (Academic Press, 1978), p. 1.
20. S. J. Orfanidis, *Optimum Signal Processing. An Introduction*, 2nd edn. (Prentice Hall, New Jersey, 1996).
21. J. H. Zar, *Biostatistical Analysis*, 4th edn. (Prentice Hall, New Jersey, 1999).