Contributions to an Advisory System for Changes Detection in Depth of Anesthesia Signals

Raquel Sebastião, Margarida M. Silva, João Gama, and Teresa Mendonça

[1] LIAAD-INESC Porto, L.A., Universidade do Porto, Rua de Ceuta, 118, 6, 4050-190 Porto, Portugal; and Departamento de Matemática, Fac. Ciências da Universidade do Porto (FCUP), Rua do Campo Alegre, 4169-007 Porto, Portugal.
raquel@liaad.up.pt
[2] Div. Systems and Control, Dep. Information Technology, Uppsala University, Box 337, SE-751 05 Uppsala, Sweden; and FCUP; and Center for Research and Development in Mathematics and Applications (CIDMA), Dep. Matemática, Campus Universitário de Santiago, 3810-193 Aveiro, Portugal
margarida.silva@fc.up.pt
[3] LIAAD-INESC Porto, L.A.; and Faculdade de Economia da Universidade do Porto, Rua Dr. Roberto Frias, 4200-464 Porto, Portugal.
jgama@fep.up.pt
[4] FCUP; and CIDMA.
tmendo@fc.up.pt

Abstract. In the clinical practice the concerns about the administration of hypnotics and analgesics for minimally invasive diagnostics and therapeutic procedures have enormously increased in recent years. The automatic detection of changes in the signals used to evaluate the depth of anesthesia is hence of foremost importance in order to decide how to adapt administered doses to patients undergoing surgical procedures. The aim of this work is to online detect changes in depth of anesthesia signals of patients undergoing general anesthesia. The performance of the proposed method is evaluated using bispectral index records. The results show that the changes detected by the proposed method are in accordance with the actions of the clinicians. This fact and the good results that were obtained support the online validation of the proposed advisory system for changes detection in depth of anesthesia signal in a real clinical environment.

Keywords: Data flow analysis, Change detection, Online learning algorithms, Anesthesia, Bispectral Index (BIS)

1 Introduction

1.1 Anesthesia Overview

Aiming to induce general anesthesia on patients undergoing surgery a combination of different anesthetic agents is commonly used. The clinicians manipulate the amount of drugs to be given to each patient in order to achieve an adequate overall anesthetic state taking into account the three main components of anesthesia: muscle relaxation, analgesia and hypnosis.

Muscle relaxation is achieved by the administration of muscle relaxants, e.g. *atracurium*, and it is quantified by the NeuroMuscular Blockade (NMB). It is measured from an evoked electromyography (EMG) at the hand of the patient by electrical stimulation of the *adductor pollicis* muscle to supramaximal train-of-four (TOF) stimulation of the ulnar nerve [2].

Analgesia and hypnosis are strongly connected in what concerns the Depth of Anesthesia (DoA). For pain relief (analgesia) the dedicated administration of opioids is usually performed. Nevertheless due to lack of reliable sensors to directly and quantitatively measure the level of analgesia, the patient's analgesic state is usually inferred by the clinician through the observation of some clinical signals such as the electroencephalogram (EEG) activity, the heart rate, and the mean blood arterial pressure. Hypnosis is the component related with unconsciousness, for which there are many available indices, e.g. Index of Consciousness (IoC) [6], Auditory Evoked Potentials (AEP) [18], Spectral Entropy (SE) [19], and Bispectral Index (BIS) [5, 7]. The BIS is the most widely used index to infer the DoA of a patient, being related with the responsiveness level and the probability of intraoperative recall [7, 13]. It is an index derived from the EEG [5] through the combination of the time domain, frequency domain and high order spectral variables [15, 7]. The BIS monitor represents the DoA in a dimensionless continuous scale ranging from 0 (equivalent to the absence of brain activity) to 97.7 (representing a fully awake and alert state). Values between 40 to 60 indicate an adequate BIS level for general anesthesia [15, 9].

1.2 Challenges in DoA monitoring and control

In the clinical practice the concerns about the administration of hypnotics and analgesics for minimally invasive diagnostic and therapeutic procedures have enormously increased in the past years. The hypnotics and analgesics usually interact with each other [11], consequently, the use of both drugs often enhances the final effect, posing difficulties to assess the correct dosage of each drug needed during the surgery. These drawbacks and the fact that the patient's response to anesthesia changes over the time-course of the surgery make the ability to monitor and control the DoA one of the main challenges of modern anesthesia.

At present, research groups working in the anesthesia field are focused on the development of advisory systems to be incorporated in automatic control platforms of DoA [10, 1, 8]. The automatic detection of changes in the DoA signals has paramount importance in the adaptation of the drug doses needed to achieve an optimum degree of comfort while avoiding undesirable side-effects [16]. Changes in DoA signals may occur due to modifications in the patient hypnosis or analgesia levels (internal factors) or due to external factors, such as intubations, incisions or other painful stimulus. The major difficulty of the success of algorithms detecting changes in BIS signals is the high presence of noise in measured data, since noise can be easily confused with initial phases of smooth drifts.

The development and analysis of change detection algorithms for the NMB play an important role in this actual challenge [17]. Owing to the characteristics of the DoA problem it is however necessary to careful redesigned the overall strategy. As a matter of fact there are different features mainly concerning the clinical sensors, the acting time and the signals that increase the difficulty to attain a robust and reliable advisory system for change detections in DoA signals.

1.3 Paper Contribution

The first contribution of this paper is the development of an algorithm to online detect changes in the BIS of patients undergoing surgical procedures. The second contribution is the offline evaluation of the performance of the proposed method using BIS records of patients subjected to abdominal surgery. The final goal of the online automatic detection of changes in the BIS measurements is to trigger an alarm in an advisory system to monitor the DoA, to help the clinician's decisions.

The paper is organized as follows: the clinical data used in this study is presented in Section 2. Section 3 describes the methodology to evaluate drifts in data, while Section 4 shows the results of this study. Finally, concluding remarks and further research are presented in Section 5.

2 Clinical Data: BIS Measurements

The BIS records used in this study were collected over the last year from patients undergoing surgery in the operating room of the Hospital Geral de Santo António (HGSA), Centro Hospitalar do Porto, Porto, Portugal.

Clinical data from 22 patients undergoing abdominal surgery was recorded, collecting both univariate sensor series (BIS, drug doses, ect) and annotations related with clinical actions (such as intubation, incision, etc) The patients were 60±15 years old, 76.75±17.74 kg and 13 female. In these cases the hypnotic propofol and the analgesic remifentanil were intravenously administrated. The DoA was manually controlled by the clinician who changed the drug doses according to clinical requirements using as reference the patient's vital signs and BIS. BIS, hemodynamic parameters and drug rates were recorded, with a frequency of $1/5s^{-1}$. The surgeries had an average duration of 144±74 minutes.

3 Methodology to Evaluate Changes in the BIS signals

In general terms, while correctly detecting changes in the data, an online drift detection algorithm must be able to forget outdated data, be robust to outliers and be single pass and run in efficiency space, allowing constant updates in time and memory. Therefore, the main challenge is a trade-off between the robustness in the presence of noise and high sensitivity to concept changes. The presence of a high level of noise and outliers constitute the main difficulties to drift detection algorithms since they may increase the number of false alarms.

In order to detect changes in the BIS records the Page Hinkley Test (PHT) was used. This test, which fullfils the previously stated requirements, is a sequential adaptation of the detection of an abrupt change in the average of a Gaussian signal [3] and is commonly used to online detect a change in signal processing [4, 12, 14]. The algorithm monitors the difference between two variables: a cumulative variable and its minimum or maximum value, depending if increases or decreases in the signal are being detected. The first variable is defined as the cumulated difference between the observed values and their current mean, where T is current time and x_T is the variable value at time T. This approach consists of running two tests in parallel, each one to detect increases or decreases in the signal. Moreover, in swift and evolving environments old data is usually less important than recent one. To address this issue, this method was enhanced with a forgetting mechanism (PHT-FM), resulting in the following tests:

For increase cases:

$$U_0 = 0$$
$$U_T = \frac{T-1}{T} U_{T-1} + (x_T - \bar{x}_T - \delta)$$
$$m_T = min(U_t, t = 1 \dots T)$$
$$PH_U = U_T - m_T$$

For decrease cases:

$$L_0 = 0$$
$$L_T = \frac{T-1}{T} L_{T-1} + (x_T - \bar{x}_T + \delta)$$
$$M_T = max(U_t, t = 1 \dots T)$$
$$PH_L = M_T - L_T$$

where the parameter δ is highly dependent of the characteristics of the signal under study. The value of this parameter is chosen to minimize false detections due to noise, taking into account the magnitude of changes that should not raise an alarm.

In these equations, the forgetting mechanism is the weight of the variables U_{T-1} and L_{T-1} in the update process. As it can be interpreted, the ratio $\frac{T-1}{T}$ increases with time, which means that the recent examples have more importance in the update process than the older ones. With this forgetting mechanism, the algorithm will be able to earlier detect both abrupt (sudden) and gradual (slow) changes.

At every instant the two PH statistics (PH_U and PH_L) are monitored and a change is reported whenever one of them is above a given threshold λ. This threshold parameter is chosen considering a tradeoff between admissible false alarm rates and delay time detections. Therefore, increasing λ the algorithm will entail fewer false alarms but might miss some true changes.

4 Results and Discussion

4.1 Offline Analysis

The PHT-FM input parameters, λ and the δ, were set to 20 and 10, respectively, taking into account the signals properties. A comparative analysis was performed to evaluate the advantage of using the forgetting mechanism. It can be noticed that with this mechanism the algorithm is able to detect the same changes as the original PHT, but earlier. This is the main advantage of the PHT-FM since reducing the delay time in detections gives more room for a clinician's decision based on that information. These comparative results are not,however, presented in this paper.

The performance of the proposed advisory system for changes detection was evaluated offline using the referred database of cases collected in HGSA, using the BIS signals as inputs for the PHT-FM. Due to difficulty of ascertain the exact time where the changes in the BIS signal occurred a delay time evaluation could not be performed. Therefore, a preliminary evaluation has been performed scoring:

- offline detections associated with a clinician's action (#DCA)
- offline detections not associated with a clinician's action (#DWCA)
- clinician's actions followed by a change that was detected by the PHT-FM (#CAD)
- clinician's actions that were not associated with a change detection (#CAWD)

These scores were obtained considering the adjustments on the drugs doses as well as some annotations related with clinical procedures (such as intubation, incision, etc). The results of this evaluation are shown in Table 1. Note that, the detections in the BIS due to the initial *bolus* of propofol and due to the end of drugs administration were not taken into account.

In spite of the high number of detection that were not associated with a clinician's action (#DWCA), this fact did not represent a major concern because clinicians might be advised and then decide an action based upon their experience. Another important result of this classification is the lower number of clinician's actions not combined with a change detection (#CAWD). This might be an evidence that the PHT-FM misses few changes (although a clinical procedure not always cause a change in the BIS signal). It should be pointed out that most the detections were classified as #DCA, which supports the online use of this algorithm as a decision support system for drug administration.

Table 1. Evaluation of the offline performance of the PHT-FM. The scores explained previously are presented for each of the twenty-two cases under study.

Case number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
#DCA	1	4	4	1	6	2	2	3	0	2	1	1	1	0	1	3	1	2	7	4	5	5
#DWCA	3	5	2	0	2	2	0	0	0	10	1	2	1	0	3	1	2	5	1	0	7	0
#CAD	0	2	1	1	3	2	1	1	0	3	1	1	0	0	3	3	1	1	5	3	6	3
#CAWD	0	2	0	0	0	1	1	0	0	1	1	1	0	0	0	0	0	0	3	1	0	0

4.2 Case studies

Figs. 1 and 2 illustrate two records of a patient undergoing a laparoscopic cholecystectomy. For each case, the top plot shows the BIS signal and the detected changes indicated by a vertical line and arrows according to increases and decreases. The changes are also marked as TP (True Positive) or FP (False Positive). A TP represents a correctly detected change while a FP (also known as type I error) indicates the error detecting a change when the signal is stable. The same representation was used in all cases in the aforementioned database.

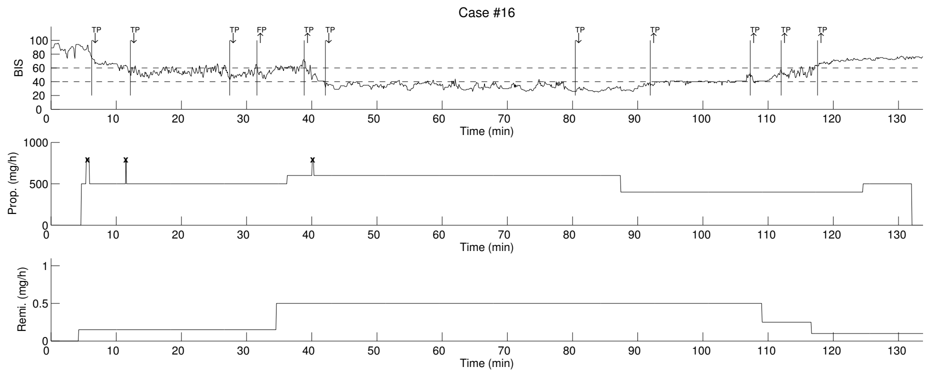

Fig. 1. Example of the detected changes in the BIS signal of case # 16 in the database. The top plot shows the BIS signal and the detected changes (indicated by a vertical line and arrows). The middle and bottom plots show the propofol and remifentanil doses (mg/h). In the middle plot, the *bolus* administration times (83 mg, 14 mg and 42 mg, respectively) are identified with a "x" symbol.

Case # 16

As it can be observed in Fig. 1, the change around minute 6, consequence of the administration of the initial *bolus* of propofol is detected by the algorithm, as expected. Since with this initial *bolus* the BIS did not decrease as much as desirable, an additional *bolus* of propofol was given around minute 12. The algorithm consistently detects the decreases of the BIS signal as a result of this propofol *bolus*, as it can also be observed later at minute 40.

In Fig. 1 it is also possible to see that the PHT-FM detects an increase and a decrease (around minute 30) that were neither consequences nor followed by any clinical action. This situation intends to illustrate the difficulties that the problem under study poses to the development of change detection algorithms, namely the false positive detections due to noise present in the BIS signals. Around minute 40 a detection of an increase in the BIS followed by the administration of a propofol *bolus* by the clinicians is noticeable. As expected, after this *bolus* the BIS decreases which is detected by the algorithm. This is one example

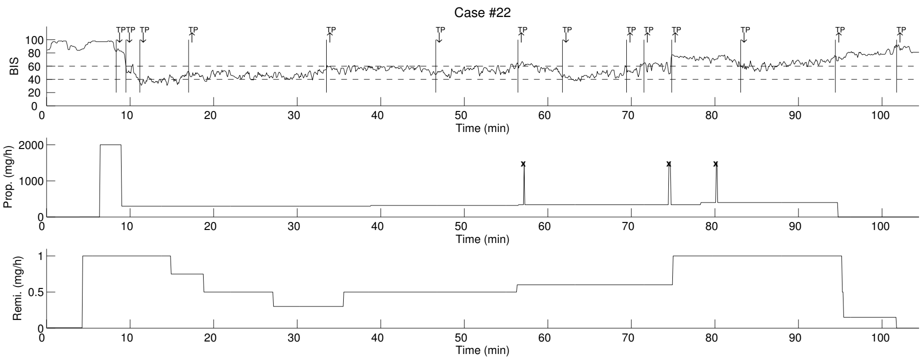

Fig. 2. Example of the detected changes in the BIS signal of case # 22 in the database. The top plot shows the BIS signal and the detected changes (indicated by a vertical line and arrows). The middle and bottom plots show the propofol and remifentanil doses. In the middle plot, the *bolus* administration times (7 mg, 21 mg and 14 mg, respectively) are identified with a "x" symbol.

where the online use of this algorithm may be advantageous: advised by the algorithm of this increase the clinician could act more promptly.

Case # 22

Fig. 2 presents a different situation. After the initial *bolus* of propofol the algorithm detected three decreases and later (around minute 15) one increase as the result of the accommodation of the BIS to this dose. In fact, between minute 15 and 30 the BIS signal remains stable around a mean value of 45. Around minute 35 the algorithm detects another increase, classified as TP due to subsequent increase in the remifentanil dose (bottom plot) in order to avoid that the BIS increases and reaches the upper level (60) of the predefined target window. One may observe that this small increase in the dosage took almost 10 minutes for the BIS to decrease as desirable (the BIS decreases around minute 45). Ten minutes later, the algorithm detects another increase in the BIS. The administration of a propofol *bolus* follows this increase in order to lower the BIS, which occurred few minutes later and was detected by the PHT-FM. A significant rise in the BIS around minute 70 is also detected by the algorithm and signaled as three increases. After this, the middle plot shows the administration of two propofol *bolus* doses in order to lower the BIS. This case exemplifies a situation where the online use of this algorithm when embedded in an advisory system could be helpful, allowing to anticipate the decision of administrating a propofol dose and avoiding the raise of BIS above the desirable threshold of 60. After these administrations the BIS recovered to the clinical reference range which was also detected by the algorithm (around minute 83). At the end of the surgery, the figure shows the BIS reaching the fully awake state induced by the end of the administration of drugs (the recovery to this state is also detected by the algorithm).

Case # 13

Fig. 3 shows a clinical case where the algorithm missed a change (indicated by a circle). The first two detected decreases are the result of the initial *bolus* of propofol. Around minute 50 it is possible to observe a false positive detection (the algorithm alarmed a change without evident existence of one). It should be noted that the noisy level of these signals poses difficulties to the drift detection algorithm and often noise can be confused with smooth drifts, alarming a drift when the signal remains stable and raising the rate of false positives. This clinical case also shows a false negative. Around minute 65, a change in the BIS can be observed, which was missed by the algorithm.

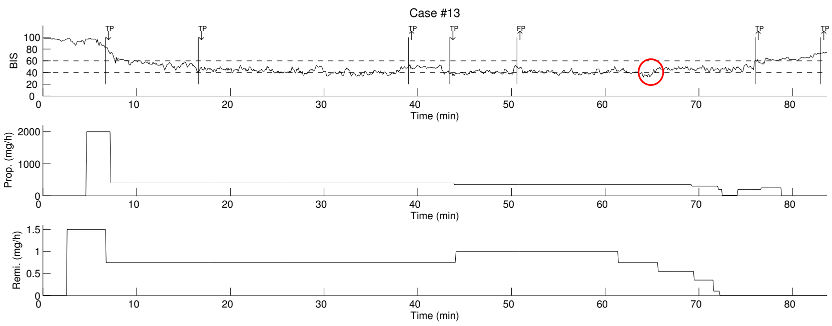

Fig. 3. Example of the detected changes in the BIS signal of case # 13 in the database. The top plot shows the BIS signal and the detected changes (indicated by a vertical line and arrows). The circle indicates a change that was missed by the algorithm. The middle and bottom plots show the propofol and remifentanil doses (mg/h).

4.3 Algorithm Performance Evaluation

To evaluate the performance of the PHT-FM, the quality metrics *Precision* and *Recall* commonly used in data mining were computed for all cases in the database:

$$Precision = \frac{TP}{TP+FP} = 87\% \qquad Recall = \frac{TP}{TP+FN} = 98\%$$

For both quality metrics, the closer to one the better the results are. The type I and type II errors were also computed with the algorithm detections for all

Table 2. Confusion matrix.

		Real		Total
		Drift	No Drift	Total
Detected	Drift	266	40	306
	No Drift	4	X	4
Total		270	40	310

clinical cases in the database. As mentioned before, a type I error is also known as a False Positive (FP). A type II error, also known as a False Negative (FN), indicates the error of not detecting a drift when in fact there exists one. Table 2 shows the above described measures. The True Negatives (TN) representing the stable points where the algorithm did not detect a drift were not assessed.

The above figures and Table 2 show that the PHT-FM identifies increasing and decreasing behaviors of the BIS, revealing the most significant changes in the signal and missing few ones, even in the presence of noise. It should be noted that, in most of the cases, the detected changes by the PHT-FM are related with an action of the clinicians. Some false positives were noticed, however those are not a major concern since clinicians might be advised and then decide based upon their experience taking into account patients' vital signals.

The performance of the algorithm has been evaluated offline on recorded BIS signals of 22 clinical cases with manual control of the propofol and remifentanil administration. This situation represents a constraint to the full evaluation of the detection algorithm. Indeed some observed clinical actions may not be associated with changes in the BIS (e.g. in Fig. 1 around minute 11, regardless the fact that the BIS is between the reference values of 40 and 60, the clinician administered a *bolus* of propofol with the *a priori* knowledge that at around minute 13 the patient will be intubated). These clinical decisions may be supported by the *a priori* knowledge of some future surgery procedures (e.g. painful stimulus). Nevertheless, the changes detected by the PHT-FM are in accordance with clinicians' assessments, in most of the cases under a great variety of sensor characteristics, supporting the feasibility of the method to be implemented online.

Those results sustain the feasibility of the proposed method as an auxiliary advisory system in surgeries to monitor the DoA.

5 Concluding Remarks and Further Research

The Page-Hinkley Test with a forgetting mechanism (PHT-FM) was implemented to offline detect changes in the behavior of BIS signals from patients undergoing general anesthesia. The developed PHT-FM algorithm consistently reveals the increasing and decreasing behaviors of the BIS signals under study.

The good performance of the algorithm when applied to real records , namely the detection of changes when the BIS signals present different behaviors and the earlier detections that allow clinicians to act more promptly, encourage an extended online clinical validation. The analysis of the obtained results supports the incorporation of this changes detection algorithm in a robust and reliable online advisory system, either for sedation or general anesthesia procedures.

As a matter of fact, the proposed approach is being used in the surgery room. With this the clinicians are able to online validate the detected changes by the PHT-FM. This consists of a strong help for the algorithm development and improvements.

It should be noted that the environment of the application and the specific features of the BIS signal, namely the high level noise present in the measurements point to further improvements of the detection algorithm.

The noise difficulty also points out to an analysis of the sensitivity of the proposed approach to the selection of the input parameters, evaluating the dependance of the number of FP and TP from both parameters' values. It must be stressed that the missing of a change point by the PHT-FM is a major concern: not being advised that a change occurred, the clinician may act later than needed. However, a misdetection is a less important occurrence: the clinician might be wrongly advised but has always the possibility to decide correctly based on patients' vital signals.

The development of a dedicate online filter to smooth the BIS signals is a future task to be addressed in order to enrich the detection algorithm's results.

Acknowledgments. The work of Raquel Sebastião and Margarida M. Silva is supported by the Portuguese Foundation for Science and Technology (FCT) under the PhD Grants SFRH/BD/41569/2007 and SFRH/BD/60973/2009, respectively. This work has been developed in the context of the Projects GALENO - Modeling and Control for Personalized Drug Administration (PTDC/SAU-BEB/103667/2008) and KDUDS - Knowledge Discovery from Ubiquitous Data Streams (PTDC/EIA-EIA/098355/2008).

The authors gratefully acknowledge Dr. Simão Esteves, from Hospital Geral de Santo António (HGSA), Centro Hospitalar do Porto, Portugal, for his participation in the collection of the clinical cases under study.

References

1. Absalom, A.R., Keyser, R. De, Struys, M.M.: Closed Loop Anesthesia: Are We Getting Close to Finding the Holy Grail? Anesthesia & Analgesia. 112(3), 516–518 (2011)
2. Alonso, H.,Mendonça, T.,Rocha, P.: A Hybrid Method for Parameter Estimation and its Application to Biomedical Systems. Computer Methods and Programs in Biomedicine. 89(2), 112–122 (2008)
3. Basseville, M., Nikiforov, I.: Detection of Abrupt Changes: Theory and Applications. Prentice-Hall Inc (1993)
4. Gama, J., Sebastião, R., Rodrigues, P.P.: Issues in Evaluation of Stream Learning Algorithms. In: 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 329–338. ACM Press, New York, NY, USA (2009)
5. Gan, T. J., et al.: Bispectral Index Monitoring Allows Faster Emergence and Improved Recovery from Propofol, Alfentanil and Nitrous Oxide Anaesthesia. Anesthesiology. 87(4), 808–15 (1997)

6. Jensen, E.W., et al.: Validation of the Index of Consciousness (IoC) During Sedation/Analgesia for Ultrasonographic Endoscopy. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 5552–5555 (2008)
7. Kaul, H. L., Bharti, N.: Monitoring Depth of Anaesthesia. Indian Journal of Anaesthesia.46(4),323–332 (2002)
8. Liu, N., et al.: Closed-Loop Coadministration of Propofol and Remifentanil Guided by Bispectral Index: A Randomized Multicenter Study. Anesthesia & Analgesia. 112(3), 546–557 (2011)
9. Luginbuhl,M., Schnider, T.W.: Detection of Awareness with the Bispectral Index: two case reports. Anesthesiology. 96(1), 241–243 (2002)
10. Mashour, G.A., et al.: A Novel Electronic Algorithm for Detecting Potentially Insufficient Anesthesia: Implications for the Prevention of Intraoperative Awareness. Journal of Clinical Monitoring and Computing.23(5), 273–277 (2009)
11. Minto, C.F., et al.: Response Surface Model for Anesthetic Drug Interactions. Anesthesiology. 92(6), 1603–1616 (2000)
12. Mouss, H., Mouss, D., Mouss, N., Sefouhi, L.: Test of Page-Hinkley, an Approach for Fault Detection in an Agro-Alimentary Production System. In: 5th Asian Control Conference, pp. 815–818. IEEE Computer Society (2004)
13. Ortolani, O. et al.: EEG Signal Processing in Anaesthesia. Use of a Neural Network Technique for Monitoring Depth of Anaesthesia. British Journal of Anaesthesia. 88(5), 644–648 (2002)
14. Page, E.S.: Continuous Inspection Schemes. Biometrika. 41(1-2), 100–115 (1954)
15. Rampil, I.J.: A Primer for EEG Signal Processing in Anesthesia. Anesthesiology. 89(4), 980–1002 (1998)
16. Selbst, S.M.: Adverse Sedation Events in Pediatrics: a Critical Incident Analysis of Contributing Factors. Pediatrics. 10(4), 864–865 (2000)
17. Silva, M.M., et al.: Total Mass TCI Driven by Parametric Estimation. In: 17th IEEE Mediterranean Conference on Control and Automation, pp. 1149–1154. IEEE Computer Society, Los Alamitos, CA, USA (2009)
18. Struys, M.M., et al.: Performance of the ARX-derived Auditory Evoked Potential Index as an Indicator of Anesthetic Depth: a Comparison with Bispectral Index and Hemodynamic Measures during Propofol Administration. Anesthesiology. 96(4), 803–816 (2002)
19. H. Viertiö-Oja, et al.: Description of the Entropy Algorithm as Applied in the Datex-Ohmeda S/5 Entropy Module. Acta Anaesthesiol Scand.. 48(2), 154–161 (2004)