EEG under anesthesia—Feature extraction with TESPAR

Vasile V. Moca [a,*,1], Bertram Scheller [b,1], Raul C. Mureşan [a,c],
Michael Daunderer [d], Gordon Pipa [c,e,f]

[a] Romanian Institute of Science and Technology, Center for Cognitive and Neural Studies (Coneural), Str. Cireşilor nr. 29, 400487 Cluj-Napoca, Romania
[b] Clinic for Anesthesiology, Johann Wolfgang Goethe University, Theodor-Stern-Kai 7, 60590 Frankfurt am Main, Germany
[c] Max Planck Institute for Brain Research, Deutschordenstraße 46, 60528 Frankfurt am Main, Germany
[d] Clinic for Anesthesiology, Ludwig Maximilians University, Nussbaumstraße 20, 80336 Munich, Germany
[e] Frankfurt Institute for Advanced Studies, Johann Wolfgang Goethe University, Max-von-Laue-Str. 1, 60438 Frankfurt am Main, Germany
[f] Massachusetts General Hospital, Dep. of Anesthesia and Critical Care, 55 Fruit Street, Boston, MA 02114, USA

ARTICLE INFO

ABSTRACT

We investigated the problem of automatic depth of anesthesia (DOA) estimation from electroencephalogram (EEG) recordings. We employed Time Encoded Signal Processing And Recognition (TESPAR), a time-domain signal processing technique, in combination with multi-layer perceptrons to identify DOA levels. The presented system learns to discriminate between five DOA classes assessed by human experts whose judgements were based on EEG mid-latency auditory evoked potentials (MLAEPs) and clinical observations. We found that our system closely mimicked the behavior of the human expert, thus proving the utility of the method. Further analyses on the features extracted by our technique indicated that information related to DOA is mostly distributed across frequency bands and that the presence of high frequencies (>80 Hz), which reflect mostly muscle activity, is beneficial for DOA detection.

© 2009 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Feature extraction techniques applied to biomedical signals have proven essential in life-science applications (automated external defibrillators, implanted pacemakers, diagnosis of epilepsia, etc.). For general anesthesia it remains a challenge to monitor the impact of anesthetics on the brain. Two recent studies showed an incidence of unwanted and primarily undetected patient awareness during general anesthesia of about 0.13% [1,2]. Since awareness and memory formation can cause severe psychological trauma [3], these studies have motivated the need for DOA monitoring devices. Substantial progress has

been made in identifying signal features that relate well to anesthetics, in a dose-dependent way, for both spontaneous electroencephalogram (EEG) and mid-latency auditory evoked potentials (MLAEPs). Consequently, monitoring devices are commercially available today [2,4–7].

An important issue in automated DOA assessment is the feature extraction technique applied to the EEG signal. The most successful commercial monitors extract a combination of features based on time- and frequency-domain (BIS: Aspect Medical Systems; Narcotrend: Monitor Technik) or entropy (Narcotrend, M-Entropy: Datex-Ohmeda) from spontaneous EEG. In addition, evoked potentials (electrical responses of

the nervous system elicited by and time-locked to external stimulation) have also kept a major role in DOA assessment [8]: certain peaks and troughs in the MLAEP decrease in amplitude and increase in latency with increasing DOA [9,10]. Other methods, extracting features in the time-domain [2,11–14] have also been developed, most based on probabilistic approaches. Such a method is the A-line ARX Index (Danmeter A/S) [15], the only commercially available monitoring device based on MLAEP.

Situations may arise in which some monitors fail to perform adequately [6,16–21]. Therefore, it has been suggested that improved DOA assessment should rely on multiple features extracted from EEG [22]. Here we propose an additional feature extraction technique, namely Time Encoded Signal Processing And Recognition (TESPAR) that is novel to the problem of EEG DOA detection. It has shown impressive performance in voice recognition and engineering applications [23–25], and being a time-domain approach, it has the possibility to capture information that is not distinguishable in the frequency-domain. We combined TESPAR with a nonlinear classification technique based on multi-layer perceptrons (MLPs), in order to validate the usefulness of TESPAR for DOA detection. The technique we introduce is not to be considered a competitor of well-established DOA monitors, but the additional features extracted by TESPAR may be useful to enhance the already established methods.

2. Materials and methods

We developed an artificial system that extracts features from the raw EEG signal using TESPAR. The features are then fed to a nonlinear MLP classifier, which is trained and tested on trials labeled by a human expert relying on a morphologically different signal (MLAEP).

2.1. Anesthesia

With the approval of the local ethics committee (Ludwig-Maximilians-University, Munich), 62 patients were enrolled in the study after having provided their written informed consent. After the induction of general anesthesia and administration of muscle relaxation, the anesthesia was maintained with a combination of hypnotics and opioids. The choice of these substances was left to the discretion of the attending anesthesiologist and the dosage was based on clinical routine. The administration of hypnotic agent was adjusted when signs of wakefulness were present and was preemptively increased before anticipated painful surgical stimulation. The MLAEP was not available to the responsible anesthesiologist (for detailed information see Appendix A.1).

2.2. Data acquisition

During the medical procedures (see also Appendix A.2) auditory stimulation was applied to the patients in the form of short clicks with a continuous repetition rate of 9.1 Hz. All the intraoperative events such as awake, induction, intubation, coughing, spontaneous breathing, response to simple or complex requests and so on were coded by keystrokes and stored along the recorded EEG. Data was recorded continuously from induction to wake-up. The EEG signal was recorded differentially between A1 and Fp2, according to the international 10/20 system [26], with a sampling rate of 4 kHz.

2.3. Data pre-processing

The amplified and recorded data, with a bandwidth of 0.5–600 Hz, were processed offline for further filtering, artifact removal (for detailed information see Appendix A.3), and rejection of power line frequency (50 Hz). Next we divided the data into 100 s long segments, recorded before and after intraoperative events. These events offered additional information to the human expert and were concomitant with actions performed on the patient (e.g. changes in the drugs administration, intubation, skin incision, etc.) or with feedback detected from the patient (e.g. blood pressure variation, tears, heart rate change, active breathing, etc.). The data from each segment were analyzed in two different ways. First, we divided the segment into short trials (110 ms long) aligned to the auditory stimulus. Segments that contained less than 600 artefact-free trials were discarded. The trials were used for the MLAEP-based classification performed by human experts. Second, segments validated previously were also analyzed in their full length (without dividing them into trials) using the TESPAR method. Subsequently, features extracted by TESPAR were used for the classification performed by MLP artificial neural networks. To further identify the importance of different frequency bands for classification, filtering was also applied on each segment, prior to feature extraction.

2.4. Human expert classification

We randomly selected 600 segments across all 62 patients that included periods with different depths of anesthesia. To manually classify the data based on the MLAEP, we computed the evoked responses by averaging 600–800 artefact-clean trials per segment. Next the MLAEPs were visually categorized into one of five classes by two human experts, each expert being unaware of the other expert's judgement. Additional information was provided by the corresponding intraoperative events (see above). The experts relied on this additional information to decide between two adjacent DOA classes. The five DOA classes were defined as follows: class 5 corresponded to an awake patient able to respond to complex verbal requests; class 4 was defined as very light anesthesia with patients able to respond to very simple requests like hand squeeze; class 3 was associated with states of sleep, in which patients do not respond to light stimuli but might react to strong ones; class 2 corresponded to the optimal anesthesia level; and class 1 was linked to too deep anesthesia, where brain activity is unnecessarily low (burst suppression).

The Observer's Assessment of Alertness/Sedation (OAA/S) [27] scale has been widely in use to develop and evaluate DOA monitoring devices with a main focus on periods when induction of anesthesia is performed or when patients return to consciousness. With the DOA scales used in this study we intended to cover the full range of clinical anesthesia. With the DOA assessment as used in our study there is a coarser resolution for the states of sedation with the DOA levels 4 and
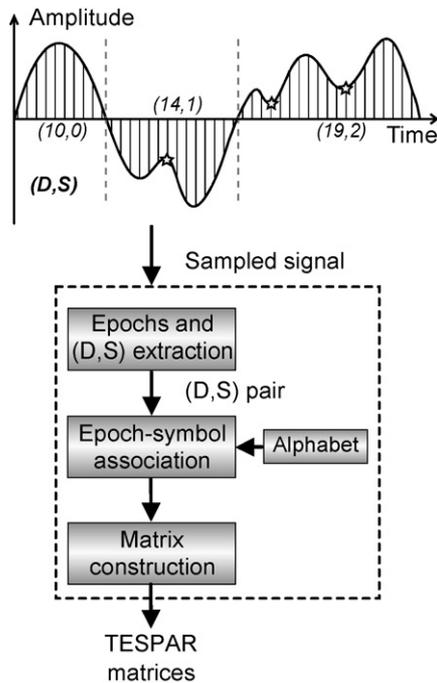
Fig. 1 – TESPAR coding diagram block. It shows how the digital signal is divided in epochs (delimited by vertical dotted lines). Each epoch is characterized by its number of samples (*D*) and its number of local minima (*S*) (we refer to the local minima of the absolute value of the waveform). The (*D,S*) pair stream (values showed for each epoch) is then transformed in a symbol stream, with the help of an alphabet that maps each pair onto a symbol (number). The symbol stream is further condensed in TESPAR A or S matrices in the final stage of the coder.

5 corresponding to OAA/S levels 2–5. On the side of deep general anesthesia the resolution of the scale used in this study is finer with the OAA/S level 1 being potentially differentiated onto DOA classes 1–3.

2.5. Feature extraction, TESPAR

First proposed by King and Glossing [25], TESPAR is a time-domain digital language for coding "band-limited" signals. The simple TESPAR model uses features that can be easily detected by visual inspection of the waveform, namely zero crossings and local extremes [28].

Since TESPAR was described in detail previously [23–25,28], we will be limited to presenting the most important aspects of TESPAR. First, the signal is split into portions situated between two adjacent zero crossings of the waveform, called epochs (Fig. 1). Each epoch is described by a pair of parameters: (*D*), called duration, characterizes the length of the epoch in samples, and (*S*), which describes the shape of the epoch by counting the number of its local minima (we used the number of local minima of the absolute value of the waveform). The pair (*D,S*) is then replaced, based on a TESPAR alphabet (codebook), by a number called a TESPAR symbol. The TESPAR alphabet is specific to each class of signals [28] (e.g. voice, EEG, and seismic vibrations), and its main purpose is

to reduce the noise affecting the epochs by assigning the same symbol to similar epochs. The alphabet approximates the distribution of epochs in the (*D,S*) plane by means of vector quantization (VQ) (e.g. Kohonen maps, K-means, Linde Buzo Gray). In our case, preliminary tests indicated that the shape of the epochs did not contain useful information; therefore, *S* was not used by the analysis. The duration *D* of an epoch gives a rough estimation of the most prominent frequency associated with the respective piece of signal. We ordered TESPAR symbols according to their corresponding duration (*D*) such that higher ranking symbols implied a longer duration, and thus, a lower dominant frequency of the epoch. By using the TESPAR alphabet, the waveform was transformed into a symbol stream which was further processed to obtain fixed-size descriptors that are called TESPAR matrices (Fig. 2).

The first matrix, called the S matrix, is a histogram of symbols counts (how many times a certain symbol is present in the signal). The second matrix, called the A matrix, is a two-dimensional matrix that counts how many times a pair of symbols, situated *L* symbols apart, appears in the signal. The *L* parameter called "lag", is usually kept constant for a certain implementation. Small or large lag values give A matrices that describe the short or long time evolution of the signal, respectively [28]. Since time is included through the lag, the A matrix provides richer information content than the S matrix. Preliminary tests (results not shown), revealed that A matrices with a lag = 1 yielded the best performance.

In addition, we applied further processing to the matrices to increase the saliency of the representation. Long epochs, which were usually more rare, were emphasized with respect to shorter, more frequent epochs. The nonlinear transformation $a_{ij} := \ln(a_{ij} + 1)$ applied on the element a_{ij} of the A matrix allowed the amplification of small values and small differences so that long epochs received increased importance. The same nonlinear transformation was applied to elements s_i of the S matrix. Before the TESPAR matrices were presented to the MLP, they were normalized to the interval [0, 1], by dividing each element in the matrix by the maximum of the matrix.

2.6. Feature mapping (classification)

To test whether TESPAR matrices contain information about DOA, we used MLP networks [29] to compute a mapping between TESPAR matrices and the five classes of DOA. MLP classifiers were chosen because they have recently recorded notable successes in anesthesia-related problems [30]. MLP topologies were constrained by the mapping problem: The size of the input layer was defined by the size of the input TESPAR matrix, the output layer had five output neurons (one for each DOA class), the size and number of hidden layers was optimized in a heuristic manner. The backpropagation algorithm was used for training [31,32].

Classification performance was further boosted [23,33,34] by employing ensembles of classifiers [34] trained by using a bagging technique (for each classifier a subset of the training set was randomly selected [35]). To avoid training bias, the number of examples was balanced, such that the training set for each class had the same size. This limited the training set
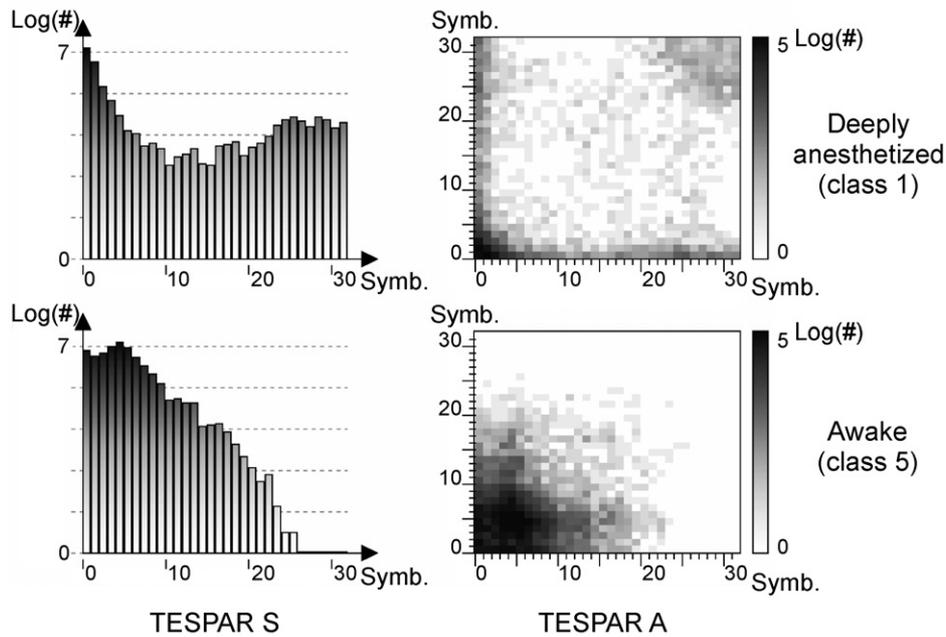
Fig. 2 – TESPAR S and A matrices ($L = 1$) for awake (class 5) and deeply anesthetized states (class 1). The S matrix counts how many times symbols appear in the data stream. The A matrix represents in a gray scale the number of occurrences for each symbol pair situated L symbols apart. The log transformation (see Section 2.5 "Feature extraction, TESPAR") has been applied to the matrices. Prominent differences can be seen in the matrices for the two extreme states (awake, deeply anesthetized).

size to the size of the smallest available training set (in our case, corresponding to class 5). After training, the MLP ensemble used a majority vote strategy to give the final classification result.

3. Results

We employed TESPAR and MLP classifiers, to extract DOA information contained in EEG segments comprising multiple trials that followed an auditory stimulus. An entire segment was considered as either training or a testing sample. The performance achieved by the artificial system (TESPAR and MLP) on the EEG segments was assessed in comparison to the classification performed by two anesthesiologists with expertise in visual classification of MLAEPs computed on the same segments.

3.1. The human classifier

We first evaluated the performance of the human experts (see Appendix A.4). For this purpose, the two anesthesiologists were asked to classify the same data on two different occasions. Anesthesiologists A and B classified 70.77% and 68.29% of the segments in the same way, respectively (see also Fig. A.1 and Table A.1). These measures of self-consistency were regarded as the highest performance that the artificial system, learning from the human expert, could achieve.

3.2. Method configuration and calibration

To calibrate the method, we have used MLPs with one hidden layer, trained and tested on TESPAR A matrices (Fig. 2). For each human expert, the validated data were half-split to obtain a training set and a testing set, and these sets were balanced according to the procedure presented above ("*Feature mapping (classification)*"). Class 5, corresponding to an awake patient, occurred most rarely (19 and 20 times for A and B human experts, respectively). Thus, the size of the training set was limited to 9 and 10 samples per class for anesthesiologist A and B, respectively.

To adjust the TESPAR alphabet size and the hidden-layer dimension, fifteen single MLPs were constructed for each parameter configuration, each MLP having a randomly assigned training subset. For each MLP the accuracy was evaluated with respect to the classification performed by one of the anesthesiologists. The median of classification accuracy over the 15 MLPs was used as the main performance indicator.

For the TESPAR alphabet size, we tested the range of 2–50 symbols. More than 15 symbols yielded a classification performance above 50%. It is known that an alphabet of about 30 symbols should be suitable for most types of signals and applications [28]. Here, the highest performance ($\sim 55\%$) was obtained for 20 and 32 symbol configurations.

Variation of the hidden-layer dimension (3, 5, 10, 15, 20, 30, 50, and 100) showed increased performance for 5 and 10 neurons. Further increase in the number of hidden-layer units resulted in a steady performance decrease. Therefore, we kept

the hidden-layer dimension as small as possible to encourage generalization.

The most suitable configuration was obtained with 32-size TESPAR alphabet in conjunction with 5 hidden units. Unless otherwise specified the following architecture was used further: input layer with 32 or 1024 units for systems using S and A matrices, respectively; 5 neurons in the hidden layer and 5 neurons in the output layer (one for each DOA class).

3.3. The artificial system performance

We first tested if there was a difference in the way that the artificial system learned one or the other human classifier. For each anesthesiologist we trained 125 MLPs on A matrices. To enhance reliability and classification performance, networks were randomly grouped in 5 ensembles (comprising 25 networks each). We computed the average and standard deviation across classification performances provided by these ensembles. For anesthesiologist A, we obtained a mean of 60.06% and a standard deviation of 1.48%. For anesthesiologist B, we obtained a mean of 45.38% and a standard deviation of 2.05%. The test indicated that the artificial system learned better from anesthesiologist A meaning that the association between segments and DOA classes for anesthesiologist A presented a structure that was easier to grasp by the artificial system. Next we considered only anesthesiologist A as a reference.

We further investigated whether MLPs with more than one hidden layer perform better than MLPs with a single hidden layer. Tests on a set of MLPs with five neurons in two hidden layers, grouped in the same configuration as above, did not indicate any increase in classification performance (average of 60.62% and a standard deviation of 1.17%) for anesthesiologist A.

In the following, we tested whether the logarithmic nonlinearity enhances the saliency of the TESPAR matrices, and whether the temporal information contained in the A matrix is beneficial. We trained four sets of 2500 MLPs, each corresponding to a different system configuration (using S or A matrices with or without the logarithmic nonlinearity). For each configuration, we analyzed the distribution of individ-

Table 1 – Self-consistency of the human expert (Y_{H-h}) and confusion matrices of the artificial system Y_{H-A} and Y_{H-S} in configuration of best 1% ensemble of $Log(AM)$ and $Log(SM)$, respectively.

	H_1	H_2	H_3	H_4	H_5
Human–Human (Y_{H-h})					
h_1	37.50%	11.72%	3.23%	0.00%	0.00%
h_2	50.00%	77.93%	22.58%	0.00%	0.00%
h_3	12.50%	9.66%	61.29%	8.77%	0.00%
h_4	0.00%	0.69%	12.90%	71.93%	50.00%
h_5	0.00%	0.00%	0.00%	19.30%	50.00%
Human– $Log(AM)$ (Y_{H-A})					
A_1	54.55%	9.59%	4.35%	0.00%	0.00%
A_2	45.45%	77.40%	23.19%	0.00%	0.00%
A_3	0.00%	11.64%	52.17%	15.52%	0.00%
A_4	0.00%	0.00%	14.49%	70.69%	60.00%
A_5	0.00%	1.37%	5.80%	13.79%	40.00%
Human– $Log(SM)$ (Y_{H-S})					
S_1	63.64%	8.22%	5.80%	0.00%	0.00%
S_2	36.36%	76.03%	23.19%	0.00%	0.00%
S_3	0.00%	14.38%	49.28%	10.34%	0.00%
S_4	0.00%	0.00%	17.39%	81.03%	60.00%
S_5	0.00%	1.37%	4.35%	8.62%	40.00%

ual MLP classification performances and the score achieved by the ensemble (Fig. 3 a). To further increase the performance we next discarded the poor MLPs [31], focusing on the best 1% of MLPs (Fig. 3b). Results indicate that the median of the performance distribution increases by 1–2% (Fig. 3 a and b) if the log nonlinearity is applied to the A or S matrices. Also, the nonlinearity increases, in various amounts, the classification performance of the ensemble (thick grey bars in Fig. 3). Ensembles, consistently achieve higher performance than the median performance over the MLPs used to construct them. Moreover, for the best 1% selection (Fig. 3b) ensembles even surpass the best network in the set. Note that A matrices offer slightly better discrimination power compared to S matrices. Finally, the highest performance was achieved by ensembles of best 1% for the $Log(SM)$ and $Log(AM)$ features, with accuracies of 69.05% and 68.03%, respectively. These results must be judged in relation to the maximum perfor-
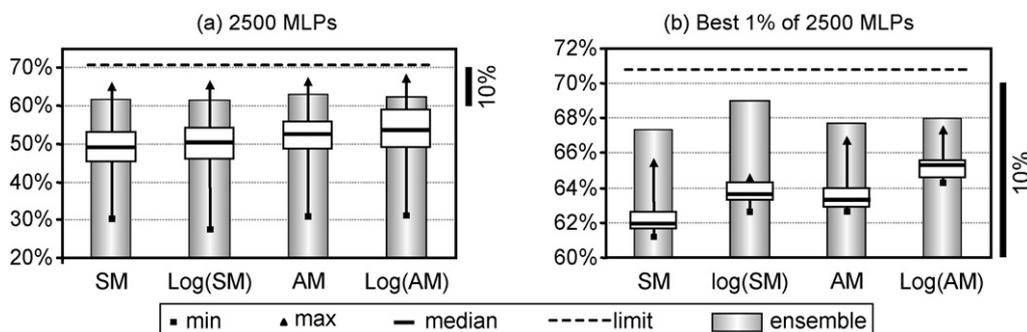
Fig. 3 – The classification performance of the four sets of 2500 MLP networks (a) and the selection of the best 1% (b). On the X-axis we indicate which TESPAR matrix is used and whether the log nonlinearity is applied or not. The dotted line represents the self-consistency of the human classifier, which is taken as the maximum achievable performance by the artificial classifier. The box plots show, for each distribution, the median, the quartiles q1 and q3, and the minimum and the maximum performances. The ensemble performance is presented through thick grey bars. Note the different scales in (a) and (b) that are emphasized by the 10% indication (thick black bars shown on the right).

mance (self-consistency) that the reference human achieved (70.77%).

We next investigated the two configurations that yielded the best performance in order to test whether the classification differed between the human expert and the artificial system. Therefore, we assessed the similarity of three confusion matrices. The first, (Y_{H-h}) (Table 1 upper pane), describes the self-consistency performed by the human expert on two different occasions. The second (Y_{H-A}) (Table. 1 center pane) and third (Y_{H-S}) (Table 1 lower pane) confusion matrices, compare the best 1% of *Log(AM)* and *Log(SM)* ensembles, respectively, to the human expert. Each confusion matrix represents the empirical probability distribution that class i was classified as class j. To estimate the similarity or dissimilarity of two probability distributions we assessed the statistical significance based on a Kolmogorov–Smirnov test, which is sensitive to any difference in the distributions (such as different means and shapes). Since the smallest p value obtained for a class was $0.85 (\gg 0.05)$ we concluded that there was no significant statistical difference between confusion matrices corresponding to the artificial system and human expert (Y_{H-A} and Y_{H-S}) and the self-consistency matrix (Y_{H-h}) of the human expert. As we expected, the errors made by the artificial system were relatively small. For the configurations presented above, the accuracy achieved was 95.24% and 96.60%, respectively, if 1 class error was tolerated (see also Fig. A.1).

3.4.   Intraoperative events and DOA levels

To gain insight in the relation between the method's prediction and OAA/S, we investigated the relations between intraoperative events and the DOA indication of the artificial system (Fig. 4). Meaningful relations were clearly visible and some examples are presented in Fig. 4. "Simple action" was detected only at DOA 4, while "Start medication" and "Coughing" were detected mostly in DOA 4–5. "Spontaneous movement" was associated rarely with DOA 2 (properly sedated) and DOA 3 but appeared more often in states with light (DOA 4) anesthesia. "Intubation" represents a very strong and disturbing stimulus and was performed during proper anesthesia (DOA 2 or sometimes DOA 1). During this procedure some patients exhibited an increase in the level of arousal (DOA 3) while others remained in the area of deep anesthesia. "Increase anesthetic" and "Reduce anesthetic" were meant to keep the patient in a properly sedated state. "Stop anesthetic" occurs usually at the end of the surgical procedures when the patient is still deeply sedated but painful stimuli are not present. Accordingly, before this event, the detected DOA level was 2 (see inset) while after the event the DOA level shifted towards more awake.

3.5.   Frequency band analysis

We tested how the information content of various frequency bands was related to the classification performance of the artificial system. The same test and training sets were used as above, with additional filtering applied to isolate or to reject frequency bands of interest such as: delta (0.5–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), beta (12–30 Hz), gamma (30–80 Hz), and high frequencies (80–600 Hz). The additional filtering was applied only on the data processed by the artificial system. The data underlying MLAEP was not filtered, such that MLAEP waveforms and, thus, the DOA segment associations remained unchanged. For each filter setting, 5 ensembles comprising 25
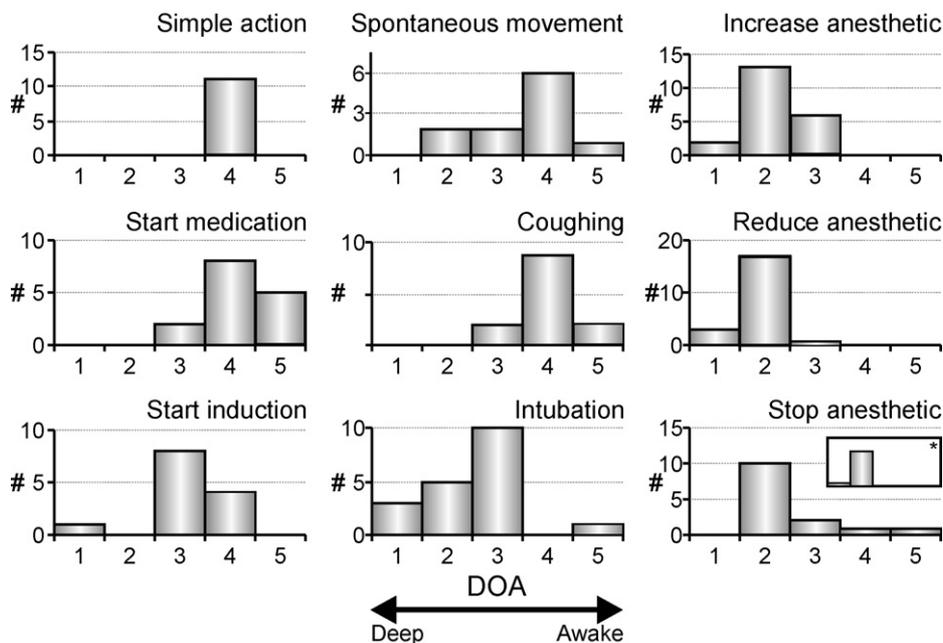
Fig. 4 – Relation between the detected DOA by the artificial system and the intraoperative events. The histograms show how the artificial system (classifier: *Log(SM)* ensemble of best 1%) assigned the DOA level to each data segment recorded immediately after a intraoperative event. The corresponding event is indicated in the top right of each pane. The inset marked with (*) shows the same plot for the data segment just before the indicated event.
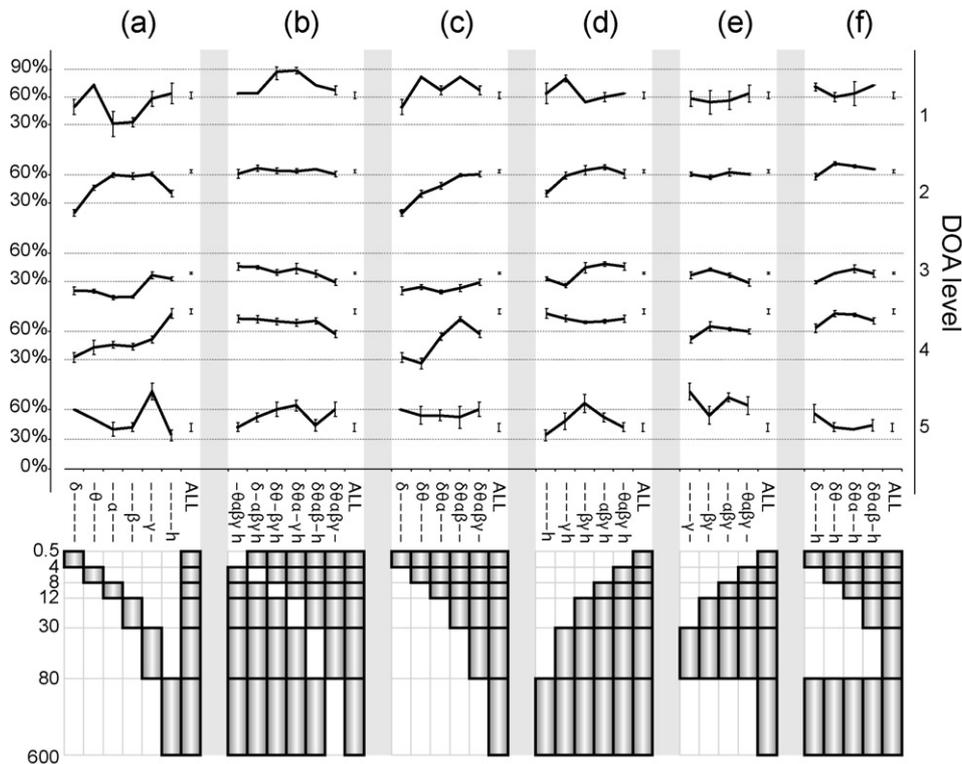
Fig. 5 – The effects of filtering on the classification performance. The bottom part shows the frequency bands preserved (filled rectangles) after filtering the EEG. For each frequency content upper traces represent the mean and the standard deviation of the classification performance for each DOA level (DOA level shown on the right). The results are grouped to emphasize: (a and b) the importance of each frequency band, (c and d) the importance of low and high frequencies, and (e and f) the importance of frequencies in the theta–gamma range.

individual MLP networks were constructed. The average and standard deviation of the 5 ensemble distributions, for each filter set and each DOA class, are presented in Fig. 5. We created sets of filters in order to: isolate one frequency band (Fig. 5a), eliminate one frequency band (Fig. 5b), preserve more and more frequency bands starting with low frequencies (Fig. 5c), preserve more and more frequency bands starting with high frequencies (Fig. 5d), and keep progressively more (Fig. 5e) or progressively less (Fig. 5f) frequency bands in the theta-gamma range. For comparison, the classification performance on the unfiltered data (Fig. 5 a–f, ALL) is shown in each case. Figure 5 illustrates a complex relation between the expression of various frequency bands and the correct detection of different DOA classes. For DOA class 1, individual alpha and beta bands are less informative (Fig. 5 a, DOA 1) and either one's absence facilitates classification when all the other bands are kept intact (Fig. 5 b, DOA 1). Delta and theta bands, when taken together, or in combination with both alpha and beta, provide a good discrimination of this DOA class (Fig. 5 c, DOA 1). Another prominent peak is interestingly obtained for a combination of gamma and high frequencies (Fig. 5 d, DOA 1). Importantly, DOA class 1 is not well discriminated when all frequency bands are lumped together, but seems to be well classified only when low (delta–theta) or high (gamma–high) frequencies are grouped (Fig. 5 c and 5 d, DOA 1), or when either alpha or beta bands are eliminated while other bands

are preserved (Fig. 5 b, DOA 1). DOA class 2 is better classified when relying on individual alpha, beta, or gamma bands than on other individual frequency bands (Fig. 5 a, DOA 2). Nonetheless, in this case, adding progressively more bands is beneficial for classification (Fig. 5 c, DOA 2). For DOA class 3, the relation between different frequency bands and classification performance is less clear, such that it is rather a combination of multiple bands, in various configurations, that can provide informative features to the classifier (Fig. 5 a–f, DOA 3). DOA class 4 is best detected when either very high (> 80 Hz) or all frequency bands are preserved (Fig. 5 a, DOA 4). Most other combinations prove to be less informative for detecting this class. Finally, DOA class 5 is best detected when only the gamma frequency band is kept (Fig. 5 a, DOA 5). Adding other frequency bands mostly impairs classification. The combination of frequency bands yielding the next best performance as compared to gamma includes the alpha, beta and gamma bands lumped together (Fig. 5 e, DOA 5).

4. Discussion

4.1. Methodological considerations

For practical reasons the following drawbacks could not be avoided. First, one of the drawbacks of this study is the low

number of training examples for the less represented conditions (mostly the deeply anesthetized state, and to a smaller degree, the awake state), which are avoided during surgery. However, the system was able to generalize, even based on just a few examples. Second, since OAA/S scale was not suitable for this study, the artificial system was evaluated with respect to the human expert, who is subject to variability and errors. This was somehow alleviated by the relations between the detected DOA and the occurrence of intraoperative events. Third, even if TESPAR does not use averaged data, still, a quite long waveform (about 100 s) is needed for reliable construction of S and A matrices; this could be problematic if rapid transitions in the sedation level should occur (e.g. when an unprecedented strong stimulus is applied to the patient). Fourth, further studies are needed before a conclusion can be formulated on which component (ongoing activity or stimulus-induced locked/non-locked activity) of the EEG signal is important for DOA assessment. These studies should make use of EEG recordings with and without auditory stimulation. Despite these drawbacks, we believe that TESPAR feature extraction proves to be valuable in DOA detection.

4.2.   *Mapping signals onto DOA*

4.2.1.   *The human expert*
The confusion matrix of the human expert is not ideal (Table 1Y_{H-h}). This is expressed through an uncertainty in the MLAEP mapping onto DOA levels. However, this uncertainty is limited to one DOA class (Fig. A.1). There are several possible reasons for that. First, the states visited by the brain passes throughout surgery might not be stationary and, thus, a MLAEP might capture aspects from more than one state. Second, the five classes defined do not have clear separation borders. Third, some of the classes chosen here are close, from the clinical point of view, to each other. The EEG signals, for these cases, have similar properties rendering MLAEP-based visual discrimination difficult (e.g. class 1 with class 2 and class 5 with class 4). This subjectivity could have been reduced if an OAA/S like observation-based scale would have been available [27,36,37]. Although, OAA/S score was not assessed we have found meaningful relations between the detected DOA level and the occurrence of intraoperative events, which are part of OAA/S assessment.

4.2.2.   *The artificial system*
MLP networks were previously used in DOA detection [30]. Other studies reported better performance if two hidden layers are used, but in our case, an extra hidden layer did not seem beneficial. However, in all cases, grouping MLP networks in ensembles made the mapping more robust, sometimes even surpassing the performance achieved by the best MLP within the ensemble.

Since we used the classification performed by the human expert (who misclassifies usually by at most one class) to train and test the artificial system, we expected at most one class erroneous classification by the latter as well. Indeed, confusion matrices between the human–human and human–system pairs were very close to each other. Thus both systems presented similar mapping patterns (Table 1). The

artificial system's performance was very close to the self-consistency of the reference human expert.

4.3.   *Features and feature extraction*

During TESPAR alphabet generation, the vector quantization eliminates the small variability present in similar epochs, endowing the system with robustness to noise. Because the symbols in the alphabet were ordered in ascending order according to their duration, each symbol could be associated with a certain frequency. Generally, as the patient went into deeper anesthesia (Fig. 2), symbols with longer duration were more frequent, and thus, lower frequencies were more prominent in the EEG. In addition, the log nonlinearity emphasizes the importance of these, more rare, long epochs. Tests (Fig. 3) showed that the log nonlinearity(see Section 2.5"Feature extraction, TESPAR") increased the separability in the DOA domain, for both A and S matrices and thus increased the classification performance. The system based on the A matrix performed, however, better than the one with the S matrix in most cases, indicating that the temporal structure of the EEG contains useful information for DOA assessment. All, these results validate the usefulness of TESPAR features for DOA detection.

Regarding the frequency content of EEG during anesthesia, we found that information related to DOA is mostly distributed across frequency bands (Fig. 5). The nonlinear classification using TESPAR and MLPs can take advantage of multiple frequencies, such that, in general, there is a combination of multiple frequency bands that yields the best detection results. This was particularly true for deeper anesthesia conditions (classes 1 and 2), for which multiple combinations of frequency bands across the spectrum are required for good classification.

In states of low sedation or wakefulness, higher frequency bands start to play a more prominent role. High frequencies ($> 80\,Hz$) reflect mainly muscle activity and are frequently considered to be an artifact. Here however, we found that their presence is beneficial. For example, frequencies >80 Hz proved useful for the correct detection of low sedation levels (class 4) and also contributed, in combination with other frequency bands, to the correct detection of other DOA classes. Hence, muscle activity may carry DOA-relevant information which should not be overlooked. Indeed muscle activity was considered useful also by the M-Entropy monitor described in Ref. [38]. Results indicate that gamma band activity (30–80 Hz) is very important for correctly identifying states of wakefulness (class 5) and that the presence of other components, especially of low frequencies, impairs the detection of such states. This is highly consistent with findings that gamma band activity correlates with visual and cognitive cortical processing, thus representing a hallmark of the awake state [39].

High frequency neuronal signals reflecting muscle or gamma band activity are usually not stimulus-locked, and hence they are not visible in MLAEP. Nevertheless, TESPAR descriptors can take advantage of non-locked activity. The good agreement between physiological correlates of wakefulness (e.g. gamma activity) and the awake class identified with MLAEP shows that the artificial classifier presented here can

learn important signal features that were not originally used by the anesthesiologists in order to classify the data.

4.4. Comparison to other methods

Other time-domain approaches to EEG signal processing, namely the zero crossing frequency [12,13] and the aperiodic analysis [14], have shown that certain points in the signal's waveform, like minima or zero crossings, contain DOA-related information. It has been suggested [11] that one major disadvantage of these two methods is that they do not take into account the small ripples of the waveform. In our case, these small ripples (the shape parameter) seemed unimportant. However, there are several differences between these two methods and TESPAR. The zero crossing frequency tries to detect the average frequency by counting how many times the signal crosses the zero level. The aperiodic analysis describes the waveform portion between two adjacent minima through its amplitude and frequency (reciprocal of its duration). Then, the frequency–amplitude pair is displayed concurrently for two frequency bands. By contrast, TESPAR takes the signal into the symbolic processing domain, where it extracts compact, fixed-size descriptors. Moreover, besides the frequency-related information (provided by the average length of the epochs), TESPAR also provides information about the temporal structure of the signal (through the A matrix).

The time-domain approach is perhaps best represented by methods based on evoked potentials (MLAEP) such as the A-line ARX Index. Evoked potentials extract stimulus-locked information that can survive an averaging procedure. Both ongoing activity and non-locked stimulus-induced activity are lost through averaging. TESPAR does not use the averaging procedure; therefore, it has the opportunity to extract more information from the signal. However, further investigations are required to answer whether the ongoing activity, the stimulus-induced activity, or perhaps a combination of these two is informative for TESPAR.

Among frequency-domain techniques, the bispectral index method is based mostly on frequency-domain signal processing (although phase, or time, information is also included in bicoherence index which is a part of the bispectral index) [11]. The Narcotrend monitor uses both time and frequency information in addition to entropy measurements [5], while M-Entropy monitor is based on entropy measurements [38]. There are several advantages of TESPAR over many of these, and especially frequency-domain, methods. First, the computational effort is at least one or two orders of magnitude smaller. Second, the only requirement imposed on the signal is to have a finite bandwidth, which is true for any signal produced by natural sources. Third, the fixed size of the TESPAR matrices, regardless of the signal's length, makes them perfect candidates for a plethora of classifiers [23–25]. Finally, since TESPAR is a time-domain method, it can detect differences in signals that look identical in the frequency-domain [28].

4.5. Concluding remarks

An intimate relation exists between the TESPAR processed EEG and DOA states of the patient undergoing general anesthesia. Indeed, the classification performance of DOA obtained with the artificial system reached very close (about 2% difference) to the mapping performed by the human expert, which indicates that DOA-relevant information is captured by TESPAR matrices. However, to reliably extract this information, we need to cope with the nontrivial relation between the descriptors and the DOA state. In order to perform the mapping, we took advantage of the human expertise, which, although biased, offered the artificial system, the possibility to unravel such a nontrivial relation. As the artificial system's performance is limited by the imperfect human expert, further studies employing more data, a more objective DOA scale, and perhaps unsupervised learning are required to asses the real value of this method and whether the ongoing and/or the stimulus-induced activity are captured by it.

TESPAR offers small, compact, fixed-size, and highly informative EEG descriptors providing, thus, a promising asset to DOA detection which could enhance already existing DOA monitors. TESPAR is perhaps most valuable in applications where light computational resources are desirable.

Conflict of interest statement

None declared.

Acknowledgments

Appendix A

A.1. Anesthesia protocol

After oral premedication with 7.5 mg midazolam 30 min prior to general anesthesia, patients were transferred to the operation theater. Patients were placed supine, and physiologic monitoring according to anesthesiologic standards, consisting of a 3 lead electrocardiogram (ECG), pulse oximetry, and oscillometric blood pressure monitoring were installed. An 18G cannula was inserted into a large forearm vein, and an infusion of isotonic saline solution was started. Then a priming dose of nondepolarizing muscle relaxant (atracurim besilate 0.01 mg/kg) and a loading dose of an opioid (left to the discretion of the attending anesthesiologist: fentanyl 0.1 mg, sufentanil 0.01 mg, alfentanil 1 mg or remifentanil continuously 0.1 mg/h) were applied i.v. and preoxygenation was started with pure oxygen via face mask in a calm environment. An equilibration period of 2 min was allowed, regardless of the opioid used. General anesthesia was induced by intravenous injection of thiopentone (4–6 mg/kg) or propofol (2–4 mg/kg)

until the loss of the eyelid reflex occurred and no response to loud verbal commands were visible. All patients were ventilated manually via face mask with pure oxygen. Muscle relaxation was induced with atracurium besilate (0.05 mg/kg) i.v. Tracheal intubation was performed 2 min after application of the muscle relaxing agent. All patients were ventilated mechanically, with minute ventilation adjusted by capnography in order to ensure normocapnia with a mixture of oxygen in air sufficient to provide an SpO_2 of greater than 97%. Balanced anesthesia was maintained with either isoflurane, sevoflure, desflurane, or propofol (the choice of substance was left to the discretion of the attending anesthesiologist), in combination with the opioid chosen for induction. The dosage of the hypnotic agent was based on clinical routine, adjusted by the occurrence of motor signs of wakefulness, while the further delivery of opioids was performed preemptively prior to periods of anticipated painful surgical stimulations, adjusted by autonomic signs of inadequate analgesia (rise in heart rate, blood pressure, sweating, and tear production). The responsible anesthesiologist was blinded towards the recorded auditory evoked potential (AEP) signal. No further muscle relaxants were applied. At the end of the surgical procedure, the delivery of all anesthetic agents was discontinued and mechanical ventilation was terminated when sufficient spontaneous breathing returned. The patients were given verbal commands repetitively every 2 min. After the first response to the verbal commands, patients' tracheas were extubated and patients were allowed to breath pure oxygen via face mask.

A.2. Data acquisition

Electrodes for AEP monitoring and headphones were applied after having inserted the iv-line. For auditory stimulation, rarefaction clicks of 98 μs duration and an intensity of 95 dB (SPL) were presented binaurally to the patient, with a continuous repetition rate of 9.1 Hz via headphones (TDH 39, Toennies/Viasys healthcare, Hoechberg, Germany). AEP were recorded by silver/silver-chloride adhesive electrodes (Neuroline 7200 00-S, Ambu/Medicotest, Denmark) placed after skin preparation with acetone, according to the international 10/20 system. Inter-electrode impedances were kept below 5 k Ω. The electrodes were connected to a preamplifier (POD, Siemens Medical, Erlangen, Germany) with short connecting cables wired to feed 4 recording channels (A1/Fp1, A2/Fp2, A1/Cz, and A2/Cz) with Fpz as common ground, according to the 10/20-system [26]. The signals were amplified and digitized (sensitivity 0.0170 μV, sampling rate 4 kHz) within the preamplifier; the digital signal was transmitted to the recording system via broadband glass fibre cables. On the recording system, the data were stored on a hard disc and, for quality control purposes, the raw EEG signal and averaged AEP epochs of 500 sweeps were displayed on a screen. All intraoperative events (awake, induction, intubation, tears, spontaneous breathing, coughing, response to complex request, and so on) were coded by keystrokes on the recording system by the investigator and stored along with the AEP-data. An interval of 2 min was allowed for the patients to rest with closed eyes for collection of baseline measurement variables. Data acquisition continued continuously throughout induction, surgical procedures, and wake-up procedure after surgical procedures had been finished. Data acquisition was stopped after the patients had regained consciousness and orientation with regard to the patient's name.

A.3. Artefact detection and removal

After rejection of 50 Hz power line frequency, we obtain EEG signals with a band of 0.5–600 Hz. Muscles, eye movements, heart electrical activity, head movement and electrical devices can generate electrical artefacts in the EEG recordings [40]. To detect the artefacts, we utilized the amplitude of the recorded signal in two frequency bands. The first band, including frequencies from 0.5 to 600 Hz, was used to detect artefacts specific to the EEG like fast drifts of the potentials. The second frequency band focuses on artefacts in the high frequency-domain, from 100 to 600 Hz, that can be caused by the surgical manipulation on the patient (e.g. cuttering). The advantage of the second high frequency band is that the biological signals have rather low amplitudes in this frequency-domain, which allows for higher sensitivity for artefact detection. For each of the two filtered signals we rejected each stimulus interval of 110 ms length for which the signal amplitude exceeded one of the two defined thresholds. The first threshold was set to 25 mV and corresponds to a limit that is not expected to be exceeded not even in recordings with very high noise. The second threshold, for each recording, corresponded to a value of four times the standard deviation of the amplitude distribution. It was introduced to detect rare and very high amplitudes likely to be non-biological. The combination of adaptive decision stage and robust feature extraction methods should allow the artificial system to cope fairly well with the remaining noise. For further selection, the data which, in the anesthesiologists' opinions, produced MLAEP-containing artefacts were excluded from this study.

A.4. The human classifier

The artificial system (TESPAR + MLP) uses the human expert (medic) to learn the association between the EEG signal and DOA. Before judging the classification accuracy of the artificial system, we must assess how well the human model classifier performs. For this purpose, in addition to the original set (600 segments), each anesthesiologist was asked to reclassify a subset of 300 randomly selected segments. Both anesthesiologists were unaware that these control segments were already presented before for classification. This allowed us to see how the same segments are classified by the same anesthesiologist, on two different occasions. Anesthesiologist A validated 284 segments, out of which 70.77% (201) segments were classified in the same way, while anesthesiologist B validated 246 segments, out of which 68.29% (168) segments were classified in the same way (Fig. A.1). We considered these scores as a measure of self-consistency for the human experts. In Table 1 the confusion matrix of anesthesiologist A (Y_{H-h}) shows in more detail the difference in classification behavior. It shows, in percentage, how many times an initial class i is classified as class j in the control classification. In the ideal case, when the human classifier is perfectly
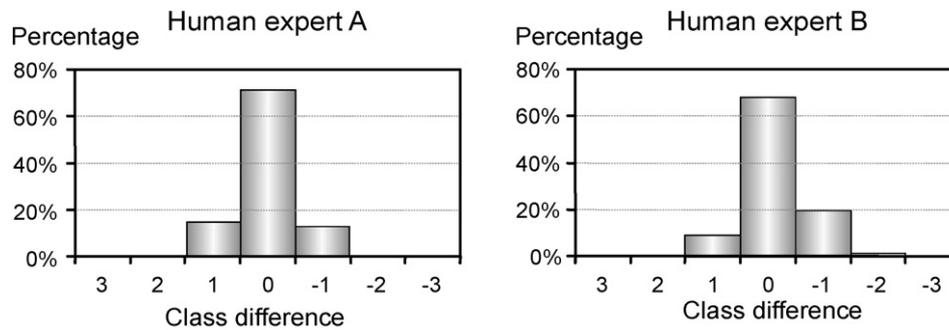
Fig. A.1 – The agreement between classifications performed on two different occasions on the same data by the same anesthesiologist. It shows how the differences between the DOA classes assigned on the two occasions are distributed. In more than 98% of the cases, the classification agreement is in within one class difference.

Table A.1 – The comparison between two human experts. It shows, in percentages, how segments classified as DOA class H_i by anesthesiologist A are classified as DOA class B_j by anesthesiologist B.					
Expert A–B (Y_{H-B})	H_1	H_2	H_3	H_4	H_5
B_1	33.33%	2.76%	1.72%	1.61%	0.00%
B_2	54.55%	81.89%	46.55%	11.29%	0.00%
B_3	12.12%	10.24%	44.83%	38.71%	23.53%
B_4	0.00%	3.54%	6.03%	41.94%	70.59%
B_5	0.00%	1.57%	0.86%	6.45%	5.88%

consistent with himself, the matrix should be 100% on the diagonal.

There has been some degree of disagreement between the two human classifiers. The difference (more than 10%) in the number of segments rejected by the anesthesiologists was an indication that they perform the classification differently. We tested to what degree they agreed with each other (Table A.1), and we found that in 61.83% of the cases they assigned the same segment to the same DOA condition. Segments that were considered to contain artifacts by any of the anesthesiologists were eliminated from this analysis. Table A.1 shows that the second anesthesiologist tends to assign deeper DOA levels in comparison to the first anesthesiologist.

REFERENCES

[1] P.S. Sebel, T.A. Bowdle, M.M. Ghoneim, I.J. Rampil, R.E. Padilla, T.J. Gan, K.B. Domino, The incidence of awareness during anesthesia: a multicenter United States study, Anesth. Analg. 99 (3) (2004) 833–839.

[2] P.S. Myles, K. Leslie, J. McNeil, A. Forbes, M.T.V. Chan, Bispectral index monitoring to prevent awareness during anaesthesia: the B-Aware randomised controlled trial, Lancet 363 (9423) (2004) 1757–1763.

[3] S. Pockett, Anesthesia and the electrophysiology of auditory consciousness, Conscious Cogn. 8 (1) (1999) 45–61.

[4] J.C. Drummond, Monitoring depth of anesthesia: with emphasis on the application of the bispectral index and the middle latency auditory evoked response to the prevention of recall, Anesthesiology 93 (3) (2000) 876–882.

[5] S. Kreuer, A. Biedler, R. Larsen, S. Altmann, W. Wilhelm, Narcotrend monitoring allows faster emergence and a reduction of drug consumption in propofol-remifentanil anesthesia, Anesthesiology 99 (1) (2003) 34–41.

[6] R.E. Anderson, J.G. Jakobsson, Entropy of EEG during anaesthetic induction: a comparative study with propofol or nitrous oxide as sole agent, Br. J. Anaesth. 92 (2) (2004) 167–170.

[7] F. Weber, H. Hollnberger, J. Weber, Electroencephalographic Narcotrend Index monitoring during procedural sedation and analgesia in children, Paediatr. Anaesth. 18 (9) (2008) 823–830.

[8] P. Martorano, E. Facco, G. Falzetti, P. Pelaia, Spectral entropy assessment with auditory evoked potential in neuroanesthesia, Clin. Neurophysiol. 118 (3) (2007) 505–512.

[9] C. Thornton, R.M. Sharpe, Evoked responses in anaesthesia, Br. J. Anaesth. 81 (1998) 771–781.

[10] E.W. Jensen, A. Nebot, P. Caminal, S.W. Henneberg, Identification of causal relations between haemodynamic variables, auditory evoked potentials and isoflurane by means of fuzzy logic, Br. J. Anaesth. 82 (1) (1999) 25–32.

[11] I.J. Rampil, A primer for EEG signal processing in anesthesia, Neuropsychobiology 89 (4) (1998) 980–1002.

[12] N. Burch, Period analysis of the EEG on a general-purpose digital computer, Ann. N. Y. Acad. Sci. 115 (1964) 827–843.

[13] F.F. Klein, A waveform analyzer applied to the human EEG, IEEE Trans. Biomed. Eng. 23 (3) (1976) 246–252.

[14] T.K. Gregory, D.C. Pettus, An electroencephalographic processing algorithm specifically intended for analysis of cerebral electrical activity, J. Clin. Monit. 2 (3) (1986) 190–197.

[15] H. Litvan, E.W. Jensen, M. Revuelta, S.W. Henneberg, P. Paniagua, J.M. Campos, P. Martnez, P. Caminal, J.M.V. Landeira, Comparison of auditory evoked potentials and the A-line ARX Index for monitoring the hypnotic level during sevoflurane and propofol induction, Acta Anaesthesiol. Scand. 46 (3) (2002) 245–251.

[16] A. Lehmann, J. Boldt, E. Thaler, S. Piper, U. Weisse, Bispectral index in patients with target-controlled or manually controlled infusion of propofol, Anesth. Analg. 95 (3) (2002) 639–644, table of contents.

[17] K. Hirota, Special cases: ketamine, nitrous oxide and xenon, Best Pract. Res. Clin. Anaesthesiol. 20 (1) (2006) 69–79.

[18] K. Kanata, S. Sakura, H. Kushizaki, T. Nakatani, Y. Saito, Effects of epidural anesthesia with 0.2% and 1% ropivacaine on predicted propofol concentrations and bispectral index values at three clinical end points, J. Clin. Anesth. 18 (6) (2006) 409–414.

[19] G. Schneider, A.W. Gelb, B. Schmeller, R. Tschakert, E. Kochs, Detection of awareness in surgical patients with EEG-based

indices-bispectral index and patient state index, Br. J. Anaesth. 91 (3) (2003) 329–335.

[20] M. Messner, U. Beese, J. Romstöck, M. Dinkel, K. Tschaikowsky, The bispectral index declines during neuromuscular block in fully awake persons, Anesth. Analg. 97 (2) (2003) 488–491.

[21] I.F. Russell, The Narcotrend 'depth of anaesthesia' monitor cannot reliably detect consciousness during general anaesthesia: an investigation using the isolated forearm technique, Br. J. Anaesth. 96 (3) (2006) 346–352.

[22] A. Schultz, M. Siedenberg, U. Grouven, T. Kneif, B. Schultz, Comparison of Narcotrend Index, Bispectral Index, spectral and entropy parameters during induction of propofol–remifentanil anaesthesia, J. Clin. Monit. Comput. 22 (2) (2008) 103–111.

[23] Y. Freund, R.E. Schapire, Experiments with a new boosting algorithm, in: L. Saitta (Ed.), ICML 1996, Morgan Kaufmann, 1996, pp. 148–156.

[24] R.A. King, T.C. Phipps, A low-power, low-complexity low-cost TESPAR-based architecture for the real-time classification of speech and other band limited signals, in: ICSPAT 2000, 2000.

[25] R.A. King, W. Glossing, Electron. Lett. 14 (15) (1978) 456–457.

[26] H.H. Jasper, The ten-twenty electrode system of the international federation, Electroencephalogr. Clin. Neurophysiol. 10 (1958) 371–375.

[27] D.A. Chernik, D. Gillings, H. Laine, J. Hendler, J.M. Silver, A.B. Davidson, E.M. Schwam, J.L. Siegel, Validity and reliability of the Observer's Assessment of Alertness/Sedation Scale: study with intravenous midazolam, J. Clin. Psychopharmacol. 10 (4) (1990) 244–251.

[28] R.A. King, T.C. Phipps, Shannon, TESPAR and approximation strategies, Comput. Security 18 (5) (1999) 445–453.

[29] P. Norvig, S. Russell, Artificial Intelligence: A Modern Approach, Prentice-Hall, 1995.

[30] C. Robert, P. Karasinski, C.D. Arreto, J.F. Gaudy, Monitoring anesthesia using neural networks: a survey, J. Clin. Monit. Comput. 17 (3–4) (2002) 259–267.

[31] S.W. Smith, Scientist and Engineer's Guide to Digital Signal Processing, California Technical Publishing, 1997.

[32] D.E. Rumelhart, G.E. Hinton, Learning representations by backpropagating errors, Nature 323 (1986) 533–536.

[33] K. Tumer, J. Ghosh, Error correlation and error reduction in ensemble classifiers, Electron. Lett. 8 (3) (1996) 385–404.

[34] G. Brown, J. Wyatt, R. Harris, X. Yao, Diversity creation methods: a survey and categorisation, Inform. Fusion 6 (1) (2005) 5–20.

[35] L. Breiman, Bagging predictors, Mach. Learn. 24 (2) (1996) 123–140.

[36] R.E. Anderson, U. Sartipy, J.G. Jakobsson, Use of conventional ECG electrodes for depth of anaesthesia monitoring using the cerebral state index: a clinical study in day surgery, Br. J. Anaesth. 98 (5) (2007) 645–648.

[37] M.S. Avidan, L. Zhang, B.A. Burnside, K.J. Finkel, A.C. Searleman, J.A. Selvidge, L. Saager, M.S. Turner, S. Rao, M. Bottros, C. Hantler, E. Jacobsohn, A.S. Evers, Anesthesia awareness and the bispectral index, N. Engl. J. Med. 358 (11) (2008) 1097–1108.

[38] H. Viertiö-Oja, V. Maja, M. S"arkelä, P. Talja, N. Tenkanen, H. Tolvanen-Laakso, M. Paloheimo, A. Vakkuri, A. Yli-Hankala, P. Meriläinen, Description of the entropy algorithm as applied in the Datex-Ohmeda S/5 Entropy Module, Acta Anaesthesiol. Scand. 48 (2) (2004) 154–161.

[39] G. Buzsáki, Rhythms of the brain, Oxford University Press, 2006.

[40] P. Anderer, S. Roberts, A. Schlögl, G. Gruber, G. Klösch, W. Herrmann, P. Rappelsberger, O. Filz, M.J. Barbanoj, G. Dorffner, B. Saletu, Artifact processing in computerized analysis of sleep EEG—a review, Neuropsychobiology 40 (3) (1999) 150–157.