



A condition-independent framework for the classification of error-related brain activity

Ioannis Kakkos¹ · Errikos M. Ventouras² · Pantelis A. Asvestas² · Irene S. Karanasiou³ · George K. Matsopoulos¹

Received: 6 April 2019 / Accepted: 26 December 2019 / Published online: 9 January 2020
© International Federation for Medical and Biological Engineering 2020

Abstract

The cognitive processing and detection of errors is important in the adaptation of the behavioral and learning processes. This brain activity is often reflected as distinct patterns of event-related potentials (ERPs) that can be employed in the detection and interpretation of the cerebral responses to erroneous stimuli. However, high-accuracy cross-condition classification is challenging due to the significant variations of the error-related ERP components (ErrPs) between complexity conditions, thus hindering the development of error recognition systems. In this study, we employed support vector machines (SVM) classification methods, based on waveform characteristics of ErrPs from different time windows, to detect correct and incorrect responses in an audio identification task with two conditions of different complexity. Since the performance of the classifiers usually depends on the salience of the features employed, a combination of the sequential forward floating feature selection (SFFS) and sequential forward feature selection (SFS) methods was implemented to detect condition-independent and condition-specific feature subsets. Our framework achieved high accuracy using a small subset of the available features both for cross- and within-condition classification, hence supporting the notion that machine learning techniques can detect hidden patterns of ErrP-based features, irrespective of task complexity while additionally elucidating complexity-related error processing variations.

Keywords EEG · ErrP · Condition complexity · Classification · Feature selection

1 Introduction

The detection of an error is the cognitive evaluation of an outcome that is considered undesired or mismatches an expected response. As such, the ability of the brain to recognize errors made during the various mental operations is an important factor for the optimization of human behavior.

Non-invasive electroencephalography (EEG) and in particular the study of event-related potentials (ERPs) elicited during incorrect actions provide new insight in the attempt to decode the complex neural mechanisms underlying error-related cognitive performance. In this regard, a negative deflection of a response-locked ERP, peaking at 40–150 ms after the commission of an error (error-related negativity, Ne, ERN) has been well-established [1, 2]. In addition, following the ERN, a positive ERP component (error positivity, PE), possibly reflecting error awareness, has been consistently reported, typically peaking at 200–500 ms after incorrect responses [2, 3]. Feedback on incorrect actions also induces a specific time-locked negative ERP (feedback-related negativity, FRN) peaking at approximately 250–300 ms after a feedback sensory stimulus [4, 5]. However, ERPs related to error monitoring do not only appear when an error occurs. Numerous studies have demonstrated an ERP component appearing after correct trials (correct-related negativity, CRN) that is similar in terms of latency and morphology to ERN and serves as an error-preventing mechanism [6, 7].

In order to unveil the cognitive processes of error monitoring, various studies that take into account different modalities

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11517-019-02116-5>) contains supplementary material, which is available to authorized users.

✉ Ioannis Kakkos
ioakakkos@gmail.com

¹ School of Electrical and Computer Engineering, National Technical University of Athens, 9, Iroon Polytechniou Str, Zografos, 15780 Athens, Greece

² Department of Biomedical Engineering, University of West Attica, Athens, Greece

³ Department of Mathematics and Engineering Sciences, Hellenic Military University, Athens, Greece

have been pursued [8–11]. To that end, EEG source localization techniques and functional magnetic resonance imaging (fMRI) have been frequently employed, revealing that the generators of negative and positive deflections such as ERN, PE, FRN, and CRN are similar in terms of topology and mostly present midline scalp distributions, suggesting that error processing is generated in the anterior cingulate cortex [10, 12–15].

The robustness of the error-related ERPs (ErrPs), a term that will subsequently be used in the present study to indicate the various ERPs stated above, has been crucial for the identification and effective analysis of error-related responses in conjunction with the overall human cognitive processes related to error monitoring. Most ErrP studies include machine learning models, which analyze the recorded EEG signals in order to recognize distinguishable patterns and construct models based on the characteristics of scalp potentials for the classification of error-related brain electrical activity as correct or incorrect [16–20].

Despite the efficient classification of ErrPs, most of these studies focus on single-task discrimination between erroneous responses. However, high-accuracy cross-task pattern recognition remains a challenge, since expansion of single-task to multi-task classification usually demonstrates poor performance [21–24]. This could be the result of training on one task and testing on another, thus being more likely to include task-specific features, reducing the overall accuracy and – in some cases – rendering the classifier unreliable in task-independent classification. In this regard, it should also be kept in mind that the extracted features vary significantly under different task conditions, while additionally the ErrP morphology has been known to exhibit significant amplitude and latency variations according to intention, psychological conditions, motivation, age, as well as among individuals [25–28].

On the contrary, no studies concerning cross-condition classification of error-related responses in regard to task difficulty have been conducted. As such, most condition-complexity error-related classification studies either focus on the modulation of ErrP components between different errors, as being affected by confidence level, error severity, etc. [29, 30], or employ machine learning techniques in different experimental paradigms of similar complexity and/or stimuli [31, 32]. For instance, Spüler and Niethammer [30] performed continuous feedback EEG classification between different types of severity errors with an average accuracy of 75%, using frequency and time-locked ERP features, suggesting that difference in classification accuracy can be attributed to task complexity. This suggestion is in line with research indicating that error-monitoring brain signals can present amplitude and latency modulations with task difficulty variations [33–35]. Endrass et al. [36] implemented a visual size

discrimination task with three difficulty conditions using principal component analysis and found decreased ERN and CRN in the highest difficulty level. Furthermore, Van der Borgh et al. [37] found significant decrements in the ERN, CRN, and partly in the PE (significant in early PE but not in late PE) during the difficult condition of a two-condition flanker task. In light of this evidence, given that the difficulty of a task may mask part of the error-processing mechanisms, demonstrating high sensitivity in their morphology under different tasks, difficulty conditions, and psychological states, the conventional predefined time windows that error-related components typically appear may be extended or overlapping.

Taking the above into consideration, the novelty of the present study concerned the investigation of error response classification in a task with conditions differing in difficulty and, more specifically, investigating whether a small number of ERP-based feature subsets can provide high cross-condition accuracy and subsequently detect condition-specific features to further increase individual condition accuracy. Moreover, our analysis included both the typical ErrP time windows and combinations of adjacent time windows, in order to reduce the effects of modifications of the ErrP signal properties due to task complexity. To investigate this hypothesis, different SVM classification methods were implemented on data collected from an auditory identification experiment with two conditions of complexity employing ERP-based time-windowed features. Taking into account the fact that the performance of classifiers is generally affected not only by the reliability and the distinctness of the features extracted but also by the number of features to be employed (in a large set some features may likely be redundant or irrelevant with respect to the classification task), we applied a feature selection (FS) framework to optimize the problem, reduce variance, and hence improve the classification performance. As such, we have employed a feature search strategy, based on the combination of sequential forward floating selection (SFFS) and sequential forward selection (SFS), to attain the optimal overall performance and at the same time assess the prominence of individual features, facilitating the investigation for the feature sets that provide high classification accuracy in relation to task difficulty. This method was capable of selecting features both common to the two conditions and specific to each condition separately, successfully discriminating between correct and incorrect responses. The high classification accuracy attained for both cross-condition and within-condition classification implies that although task difficulty might affect the characteristics of ERP components reflecting error processing, machine learning methods can efficiently detect distinct ErrP differentiations between correct and incorrect decisions.

2 Materials and methods

2.1 Subjects

The EEG data used in the present study were collected as part of a previous research [38]. In particular, the data were acquired from 14 healthy individuals (8 males and 6 females) with a mean age 26.6 ± 2.9 years, who performed an auditory identification task in two sessions, each under a different condition of complexity. All subjects were right-handed with normal hearing – measured by pure-tone audiogram (thresholds < 15 dB HL) – and no history of hearing problems. Prior to the experiment, all subjects performed an acoustic pre-test to assess their hearing ability in the frequency ranges of interest to the experiment, where they were able to successfully discriminate the tones presented. Furthermore, informed consent was obtained from all participants.

2.2 Experimental design

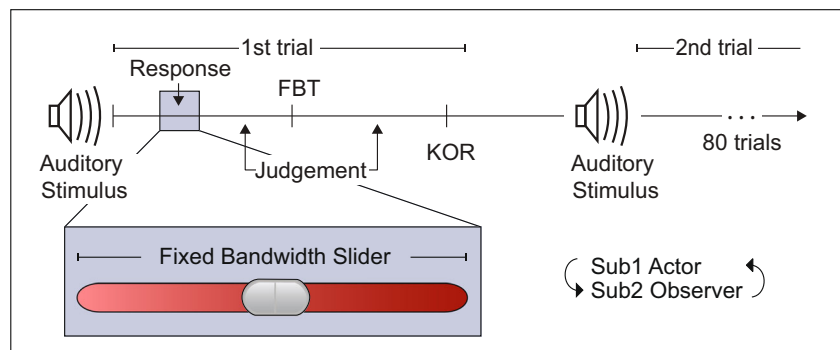
The full details of the experimental design are given in [38] and are exposed in the present section to the extent needed for the reader’s convenience (Fig. 1). Initially, subjects were divided into seven dyads, with each dyad undergoing two sessions of an auditory identification experiment under two complexity conditions. Both sessions were performed on the same date and required the determination of the specific frequencies corresponding to the acoustic stimuli. Each session consisted of 80 trials, and during each trial, the dyad members assumed actor-observer roles, switching roles among trials, thus resulting in 40 trials for each individual when participating as an actor or as an observer, respectively. Participants sat

opposite while being screened from each other and had computer screens in front of them, displaying a slider and a cursor.

At the beginning of each trial (operating phase), the stimulus was provided to both members through headphones as a 1-s duration tone, randomly selected from a block of four frequency ranges with a fixed bandwidth of 400 Hz: 200–600 Hz, 620–1020 Hz, 1040–1440 Hz, and 1460–1860 Hz. Then, the actor was asked to match the frequency of the stimulus tone via a gamepad by positioning a cursor in a slider bar appearing in both participants’ computer screens. The slider represented the frequency range, while the position of the cursor corresponded to a specific tone within this range. Participants were not aware of the band of the frequency range in which they had to place the cursor and neither the actor nor the observer could hear the sound corresponding to the position chosen during the gamepad handling. The end of the operating phase was marked by the non-movement of the gamepad for 0.5 s.

Following the operating phase, the two participants were asked to judge the correctness of the position chosen by the actor using a two-button controller (correct/incorrect). After the first judgment, the tone corresponding to the position chosen by the actor was provided to the participants (feedback tone, FBT), who were then asked to judge for a second time whether the tone corresponding to the position chosen by the actor was the same as the original. The disclosure on whether the position chosen was right or wrong was made via a “knowledge-of-results” tone (KOR). The KOR tone was either a 500-Hz tone, when the position selected was correct, or a 3-kHz tone, when the position selected was incorrect. In addition, both participants would hear the word “correct” or “incorrect” depending on the matching or not of the stimulus with the position chosen.

Fig. 1 The design of the experimental protocol. In a single trial, both subjects heard the same auditory stimulus (tone). The subject being the actor had to match the frequency of the stimulus tone by positioning a cursor in a slider appearing in both participants’ computer screens (response). Then a judgment was asked to be made concerning the correctness of the response, and a first feedback tone (FBT) was presented to the subjects. Next a judgment was again asked to be made by the subjects, and a second, definitive feedback tone (“knowledge-of-results,” KOR) was presented. Subjects alternated in actor and observer roles in successive trials



As the participants interchanged their roles as actors and observers between trials, the complexity difference between the two conditions was based on whether the frequency ranges of the acoustic stimuli belonged to the same (“easy” condition, *Joint1*) or different (“difficult” condition, *Joint2*) frequency bands. Under condition *Joint1*, acoustic stimuli of the same frequency range were presented to both participants, while in condition *Joint2*, the stimulus presented to each participant as an actor differed in terms of frequency range from the stimulus presented to his/her partner, when the partner was the actor. In this respect, observers in condition *Joint1* were expected to be more efficient in correctly matching the stimulus sound when they became actors, since they could mentally map the frequency range of the slider bar while observing their partner-actor in the previous trial. In contrast, during condition *Joint2*, by observing the actor in previous trials, individuals could not use the same mental map based on the frequency range which their partner acted on, while additionally they could be mentally disoriented when it was their turn to assign the cursor to the stimulus tone. Hence, the dissimilarity of the frequency tone would hinder the identification process and thus increase the complexity of the task, making it more challenging to identify its correct position within the frequency range employed. Indeed, this affected the individuals’ performance, as was indicated by the behavioral analysis in [38].

Within the experimental design, both FBT and KOR can be considered as feedbacks, FBT being the first-level feedback, providing indirect information for the actor’s response, and KOR being the second-level feedback, providing the unambiguous information on the correctness of the actor’s initial selection. However, since FBT was considered the first feedback for the actor’s response, it might be assumed that it elicits a cognitive response temporally closer to the action, compared to the one elicited by KOR. Therefore, only actors’ FBT ERPs were investigated in this study.

2.3 Data acquisition and pre-processing

Electrophysiological recordings were performed simultaneously for both participants, alternating between actors and observers. EEG was recorded continuously using two different recording systems, each with a 32-channel electrode cap (Biosemi, ActiTwo System), the international 10–20 EEG system. The electrodes used were Fp1, AF3, F7, F3, FC1, FC5, T7, C3, CP1, CP5, P7, P3, Pz, PO3, O1, Oz, O2, PO4, P4, P8, CP6, CP2, C4, T8, FC6, FC2, F4, F8, AF4, Fp2, Fz, and Cz (Fig. 2A). Additionally, horizontal and vertical electrooculograms (EOG) were recorded. For interference elimination, the experiment was conducted in a Faraday room,

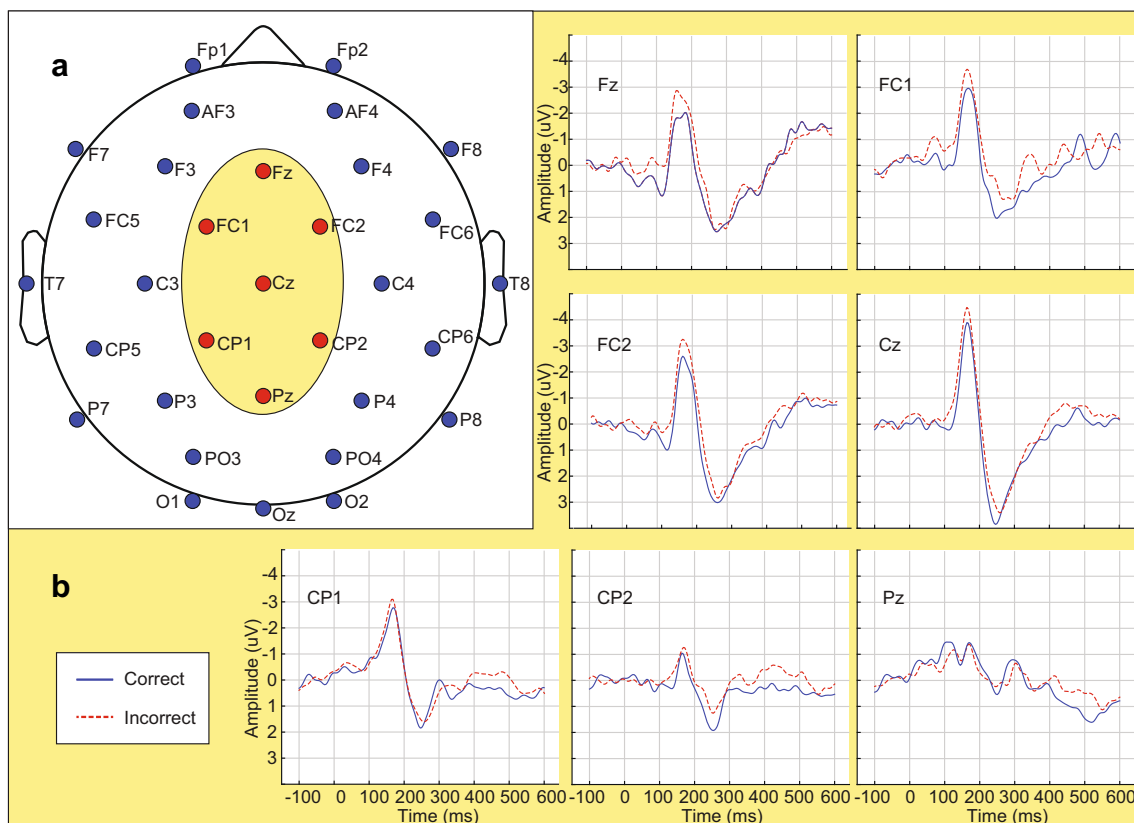


Fig. 2 (a) Electrodes used for the EEG recording. The ellipse encircles the electrodes whose recordings were employed in the feature extraction process. (b) Average ERPs across all subjects (both conditions merged) for the two classes of response

while optical receiver for trigger inputs was also used in both subjects, and electrode cables were bundled.

Recorded signals were digitized at 256 Hz and filtered offline by applying a low-pass and a high-pass Chebyshev filter with cut-off frequencies of 35 and 0.05 Hz, respectively, as well as re-referenced to the average of the electrode recordings and de-trended. Subsequently, EEG signals were segmented into ERP epochs with a duration of 2.5 s (0.5 s before and 2 s after the FBT), resulting in $40 \times 14 \times 2 = 1120$ trials acquired from the total of 40 trials for each of the 14 participants and for the 2 complexity conditions. After segmentation, each trial was baseline-adjusted relative to a 100-ms pre-stimulus baseline, and trials with ocular artifacts were manually removed. It is noted that due to significant artifact contamination, measurements of 1 dyad were excluded from subsequent analysis, leaving 12 subjects for further processing.

2.4 Definition of correct and incorrect responses

Since two tones, close in terms of frequency, can be commonly misinterpreted as the same sound, it can be assumed that similar feedback and stimulus tones may not elicit error cognition. Therefore, to evaluate the proximity of the response and stimulus tones and label the actors' responses as correct or incorrect, the individuals' ability of perceiving and discriminating between different tones was taken into account. To that end, the distinguishability of auditory perception was quantified through the psychoacoustic function of equivalent rectangular bandwidth (ERB) [39]. This function gives an approximation of the frequency range in which auditory stimuli are considered identical by modeling the filters of human hearing as rectangular band-pass filters determined as a function of a central frequency (Fig. 3). ERB was calculated by the following formula (where B_e is the bandwidth of the filter

in Hz and f is the central frequency) (presented as the stimulus tone) of the filter in Hz:

$$B_e = 6.23 \cdot 10^{-6} f^2 + 9.339 \cdot 10^{-2} f + 28.52 \tag{1}$$

Since the ERB is not a linear function (although appearing to be so in low frequencies, Fig. 3B), the use of a specific pre-defined criterion, such as the ratio f/B_e , for the definition of correct and incorrect answers could render the discrimination between the different responses ineffective. Therefore, for each trial, the individual's response was compared to the stimulus tone plus/minus the ERB bandwidth. If the response was within this range, the trial would be considered correct (Fig. 3A), otherwise it would be regarded as erroneous.

Due to the different number of response ERPs per subject and condition and on the basis that ErrPs are subject-sensitive, the pre-processed ERPs were averaged per subject and class (correct/incorrect) of the responses given, including both conditions. Specifically, for each of the 32 electrode positions, the mean ERPs were calculated for the 2 conditions and for the 12 actors according to the class of their responses (correct/incorrect), aiming to address the problem of the imbalanced classes that would impair FS and classification. Therefore, from the available data, $12 \times 2 \times 32 = 768$ FBT ERP recordings were used corresponding to correct responses, as well as $12 \times 2 \times 32 = 768$ FBT ERP recordings corresponding to incorrect responses. In Fig. 2B, we present the average across all subjects and conditions for the two classes of responses for the electrodes employed in the subsequent analysis.

2.5 Feature extraction

Although the inclusion of temporal electrode positions might provide a better insight concerning auditory cognition, central

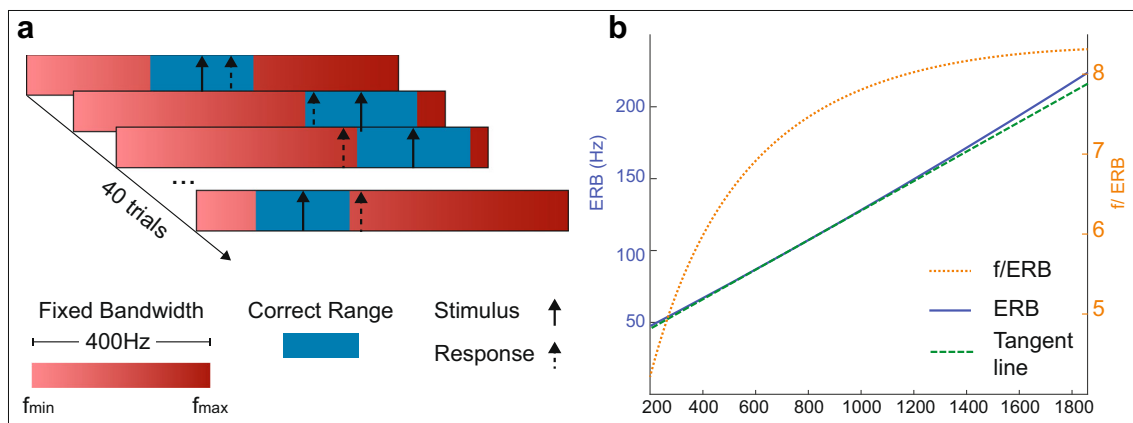


Fig. 3 (a) Definition of the correctness of each response using the equivalent rectangular bandwidth (ERB). On each trial, the stimulus (solid-line arrow) was randomly selected within the fixed frequency band (graded bar, lower left), while the response (dashed-line arrow) was determined as correct or incorrect, based on whether or not it was contained

inside the ERB range (solid-filled bar). (b) ERB as a function of the stimulus frequency (solid line), the tangent line for a random frequency (dashed line) indicating the non-linearity of the ERB function and the f/ERB ratio (dotted line)

regions of the scalp are more relevant to error processing, as stated in the introduction. Since our goal is the investigation and subsequent classification of error-related processing with regard to complexity modifications, we excluded locations not strongly related to error processing, as indicated by other studies [21, 28]. To that end, features were extracted only from the Cz electrode and from six additional electrodes (Fig. 2A). Their selection was determined according to their position relative to the Cz electrode, comprising the two midline electrodes adjacent to Cz (Fz and Pz) and the four non-midline electrodes that are closer to Cz (FC1, CP1, CP2, FC2).

Starting from the presentation of the FBT (0 ms), features were extracted from five time windows: Time window 1 (tw_1) starting at 0 ms and ending at 125 ms, time window 2 (tw_2) starting at 125 ms and ending at 220 ms, time window 3 (tw_3) starting at 220 ms and ending at 300 ms, time window 4 (tw_4) starting at 300 ms and ending at 400 ms, and time window 5 (tw_5) starting at 0 ms and ending at 600 ms. The time windows were selected in order to better isolate ErrP components of interest, as indicated by the literature presented in the introduction. Furthermore, the inclusion of the whole duration of the after stimulus ERP recording (tw_5) might provide useful features that could otherwise go unnoticed when extracting the features from the separate (small-duration) time windows.

The features calculated for each electrode position and each of the time windows were based on latency and shape characteristics describing ErrPs [16, 40] (Fig. 4) and consist of the following:

- MaxA: The maximum of the ERP signal, corresponding to the highest amplitude value for each time window
- MinA: The minimum of the ERP signal, corresponding to the lowest amplitude value for each time window
- MaxT: The latency of the maximum value, corresponding to the time MaxA occurred for each time window
- MinT: The latency of the minimum value, corresponding to the time MinA occurred for each time window

- AUC: The area under the ERP curve, estimated by calculating the ERP integral over the corresponding time window

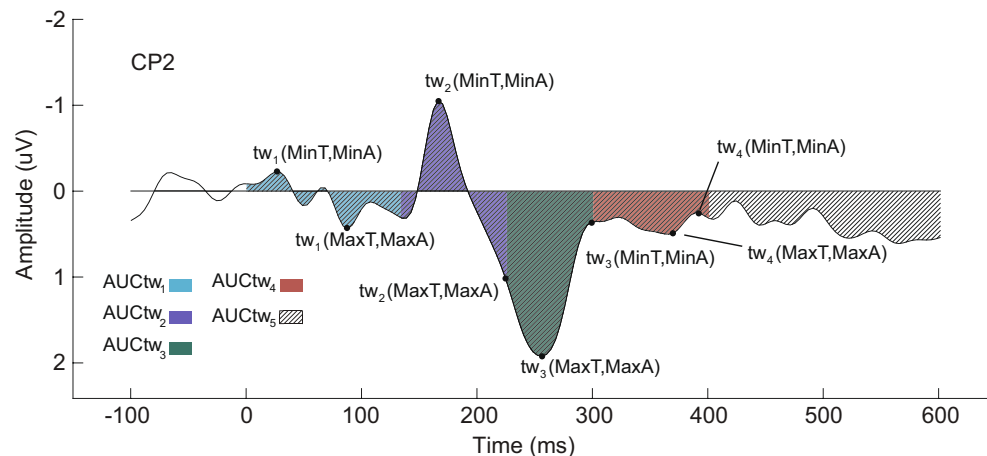
Hence, from each averaged ERP, five features were calculated for each of the five time windows and each of the seven electrode positions, resulting in $7 \times 5 \times 5 = 175$ features.

2.6 Feature selection and classification

In the present study, classification was used to discriminate between correct and incorrect responses of actors. More specifically, SVM classifiers were adopted with different configurations regarding the learning methods and kernel functions [41, 42]. The SVM framework applied included a sequential minimal optimization (SMO), a least squares (LS), and a quadratic programming (QP) SVM learning method while additionally employing linear kernel ($K(\vec{x}, \vec{z}) = (\vec{x}^T \vec{z})$), radial basis function (rbf) ($K(\vec{x}, \vec{z}) = e^{-\gamma \|\vec{x} - \vec{z}\|^2}$, $\gamma = 0.055, 0.08, 0.125, 0.22, 0.5$), quadratic ($K(\vec{x}, \vec{z}) = (c + \vec{x}^T \vec{z})^d$, $c = 1, d = 2$), multi-layer perceptron (mlp) ($K(\vec{x}, \vec{z}) = \tanh(k \vec{x}^T \vec{z} + d)$, $k = 1, d = -1$), and polynomial ($K(\vec{x}, \vec{z}) = (c + \vec{x}^T \vec{z})^d$, $c = 1, d = 3$) kernel functions. For each classification technique, the overall classification accuracy, sensitivity, and specificity were computed, which are defined as follows:

The overall classification accuracy is defined as the ratio of the correctly classified responses, i.e., the number of true positives (correct responses classified) plus the number of the true negatives (incorrect responses classified), to the total number

Fig. 4 A representation of the features extracted for time windows (tw) 1 to 5 of electrode CP2. For each tw , the time and amplitude of the minimum and maximum values of the signal, as well as the area under the curve (AUC), were calculated. The time and amplitude features of tw_5 refer to the global minimum and maximum, from 0 ms to 600 ms, and AUC is computed using the total area, indicated by the striped pattern



of responses:

overall classification accuracy

$$= \frac{\sum \text{True Positive} + \sum \text{True Negative}}{\text{Total number of cases}} \quad (2)$$

Sensitivity is the ratio of the correct responses that are classified as such, to the total number of correct responses:

$$\text{sensitivity} = \frac{\sum \text{True Positives}}{\text{Total number of Positives}} \quad (3)$$

Specificity is the ratio of the incorrect responses that are classified as such, to the total number of incorrect responses:

$$\text{specificity} = \frac{\sum \text{True Negatives}}{\text{Total number of Negatives}} \quad (4)$$

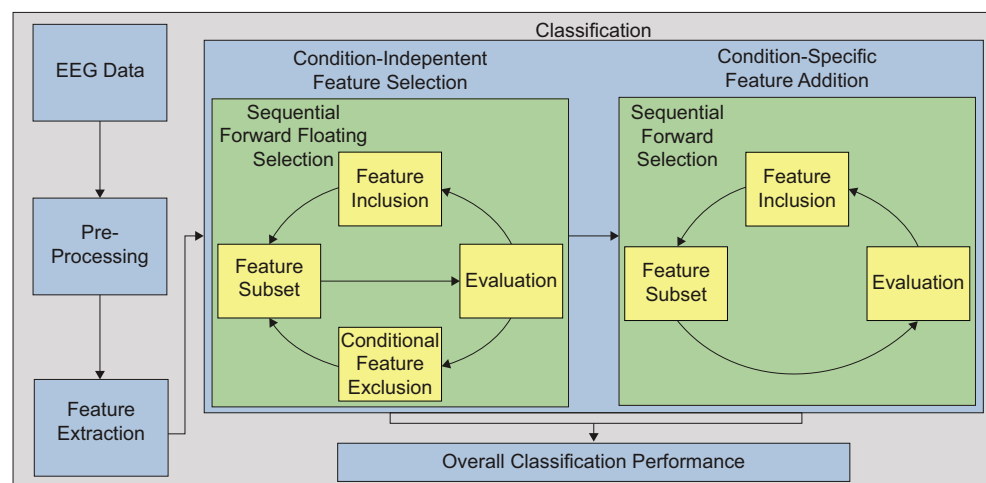
As stated in the introduction section, our main goal was to detect cross-condition high-accuracy classification feature subsets and then, on top of those features, to identify additional complexity-specific ErrP features that would improve the classification of the individual difficulty levels. To that end, we first implemented FS and classification on 12 subjects for both conditions and response classes concurrently ($12 \times 2 \times 2 = 48$ instances) reaching an FS condition-independent subset, and, subsequently, starting from that subset, we obtained task-specific features further increasing the performance on each individual condition ($12 \times 2 = 24$ instances, i.e., 12 subjects for both response classes). As a general methodological procedure, FS was applied for the purpose of examining whether specific subsets of features provide better classification performance compared to the full feature set, as well as in order to eliminate features that could carry redundant and/or unnecessary information. In this direction, the FS and

classification processes were implemented individually for the five time windows, as well as for two-window combinations: tw_1 and tw_2 ($tw_{1,2}$), tw_2 and tw_3 ($tw_{2,3}$), and tw_3 and tw_4 ($tw_{3,4}$). Overlapping windows were avoided, since they might include features from multiple components and thus mask the individual ErrP contribution to the classification process, as well as to further investigate discriminative characteristics of the ERP components and determine whether using features from components belonging to adjacent time windows might improve classification.

For the identification of the optimal condition-independent feature subset, sequential forward floating search (SFFS) [43] was applied to all extracted features that were previously extracted. SFFS is thought to satisfactorily cope with the nesting problem found in other FS methods [43, 44] and consists of an iterative repetition of three steps: inclusion, conditional exclusion, and continuation of conditional exclusion. Starting from a null set, the SFFS algorithm selects and adds into the set the most significant feature in terms of classification accuracy through an exhaustive search. Then, the new most significant feature – with respect to the existing feature subset – is included. Provided that the resulting subset will include at least two features, the least significant feature of the subset is excluded, and the new subset accuracy is estimated. Should the least significant feature be the one just added, the feature is kept in the subset and a new inclusion is made. Otherwise, a new exclusion is made with the condition that the accuracy of the new subset is better than the one found so far with the same size feature subset. This process is conducted for all features in the subset until these conditions cease to be satisfied. Subsequently, a new inclusion is conducted, and the three-step procedure is repeated until no further improvement can be by modifications of the feature set.

To ensure that the output feature set would be representative of both conditions concurrently and no bias toward a specific condition would be introduced, every SFFS step was evaluated as the average of the corresponding feature

Fig. 5 The workflow of the feature selection methodology



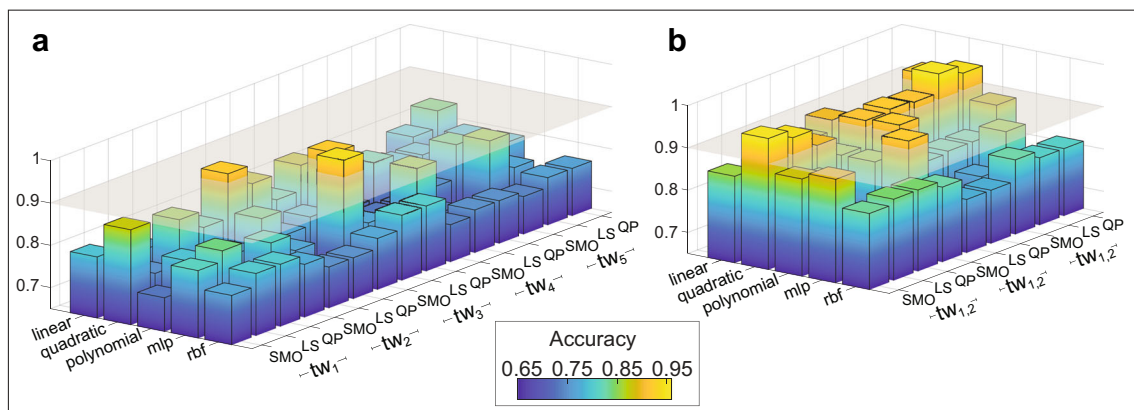


Fig. 6 Mean classification accuracy of *Joint1* and *Joint2* for all the methods employed, (a) for the individual time windows and (b) for the time windows combinations. The elevated plane represents the threshold

set accuracies of both conditions (using concurrently *Joint1* condition data and *Joint2* condition data). In this manner, the final set produced by the SFFS is deemed to represent features that best classify responses as correct or incorrect, irrespective of the task difficulty. The above procedure was repeated until SFFS concluded producing the overall values of the classification accuracies and feature subsets as the output.

Upon selection of the optimal feature subset, a sequential forward selection (SFS) method was applied in the two conditions separately [44]. Specifically, SFS started from the optimal feature subset provided by the SFFS procedure and repeatedly included the most significant feature with respect to the preceding feature subset through exhaustive search, until the classifier accuracy could not improve. To mitigate the nesting problems occurring from the greedy nature of SFS, the implemented algorithm considered a two-feature addition to the feature subset if accuracy did not increase just by a single feature addition, provided that each of the two single features would not reduce the accuracy of the modified feature subset. The termination of SFS for each condition was expected to provide the additional features that improve the

of the performance evaluation criterion of task-specific average accuracy larger than 0.9. SMO, sequential minimal optimization; LS, least squares; QP, quadratic programming SVM learning methods

classification for the specific difficulty level of each condition separately.

The above procedure (Fig. 5) was repeated for each classifier configuration, while the objective function of classification accuracy allowed for concurrent evaluation of the FS processes as well as the various classification algorithms. For the purpose of training and testing, a leave-one-out cross-validation procedure was implemented in every step of the SFFS and SFS. This procedure was adopted due to the limited data available and involves using a single instance from the original data as the testing set and the remaining data as the training set. This process is repeated, selecting a different instance each time, until all responses are used for testing once. Typically, leave-one-out cross-validation procedures provide a reliable generalization framework, approximating the actual performance of the classifiers better than other cross-validation approaches and avoiding overtraining [45, 46]. In addition, to ensure that FS introduced no bias and to assess the statistical significance of the computed accuracy values, 1000 runs of permutation tests were carried out by performing classification on randomized class labels, thus obtaining an empirical distribution of accuracy.

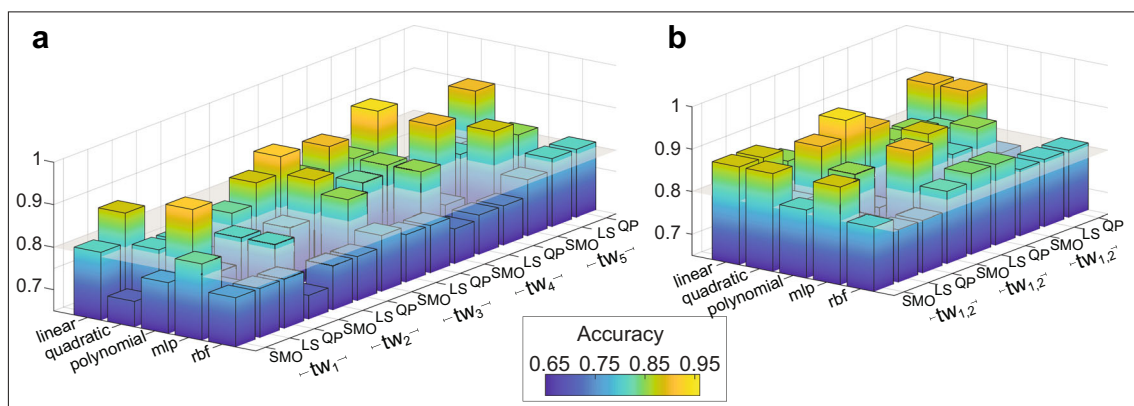


Fig. 7 Cross-condition classification accuracy for all the methods employed, (a) for the individual time windows and (b) for the time windows combinations. The elevated plane represents the threshold of

the performance evaluation criteria of cross-condition accuracy larger than 0.8. SMO, sequential minimal optimization; LS, least squares; QP, quadratic programming SVM learning methods

Table 1 Overall classification accuracy results

Time window	Classifier	Overall classification accuracy				Number of features		
		Cross-condition	Joint1	Joint2	Task-specific average	Common	Joint1	Joint2
tw_2	LS-quadratic	0.83**	0.88*	0.96**	0.92	5	5 + 2	5 + 3
tw_3	SMO-mlp	0.85**	0.96**	0.92*	0.94	5	5 + 5	5 + 4
$tw_{1,2}$	SMO-quadratic	0.88**	0.92**	1.00**	0.96	7	7 + 1	7 + 5
$tw_{2,3}$	LS-quadratic	0.90**	1.00**	0.83*	0.92	9	9 + 10	9 + 8
$tw_{2,3}$	QP-quadratic	0.85**	0.92**	0.92**	0.92	9	9 + 3	9 + 8
$tw_{2,3}$	SMO-mlp	0.90**	0.92**	0.92*	0.92	9	9 + 6	9 + 7
$tw_{3,4}$	QP-linear	0.90**	0.92**	0.92**	0.92	12	12 + 5	12 + 10
$tw_{3,4}$	SMO-quadratic	0.85**	1.00**	0.83*	0.92	6	6 + 3	6 + 1
$tw_{3,4}$	LS-quadratic	0.83**	0.92**	1.00**	0.96	2	2 + 4	2 + 3
$tw_{3,4}$	QP-quadratic	0.90**	0.96**	0.92**	0.94	7	7 + 16	7 + 2

Asterisks mark the level of significance * $p < 0.01$; ** $p < 0.001$

3 Results

The overall classification accuracy results following the FS method are presented in Figs. 6 and 7. The accuracy values on which the classifiers were evaluated were the cross-condition classification accuracy achieved by SFFS, the task-specific classification accuracy achieved by SFS when applied to the data of the two conditions separately (*Joint1* and *Joint2*), and their average value (task-specific average) which was calculated as the mean value of the task-specific accuracies of *Joint1* and *Joint2*. Due to the large number of the different methods and kernels employed, only the cases that passed a performance evaluation criterion of cross-condition accuracy larger than 0.8 and a task-specific average larger than 0.9 are further analyzed. For these cases, the corresponding results for classification accuracy are given in Table 1, while for sensitivity and specificity, they are given in Table 2. In these cases, only a small fraction of the total number

of features was selected after both SFFS and SFS were applied, with a mean feature number of 12.6 and 12.2 for *Joint1* and *Joint2* conditions, respectively.

From Tables 1 and 2, it can be deduced that FS, using the SVM classifier with rbf kernel, did not produce results that passed the performance evaluation criterion, in contrast to linear, quadratic, and mlp kernels. Furthermore, in most cases, the performance evaluation criterion was met for features extracted from combination of two time windows. On the other hand, the extended time window tw_5 that included the ERP recordings from 0 to 600 ms did not produce results meeting the criterion. Additionally, classification accuracy equals to 1 was reached for SFS in four cases, using quadratic kernels. In those cases, the optimal performance attained by task-specific average was 0.96 for two cases. The very low p values of the permutation tests, as well as the small feature subsets compared to the overall number of features, suggest that the classifiers adopted were

Table 2 Sensitivity and specificity results

Time window	Classifier	Sensitivity			Specificity		
		Cross-condition	Joint1	Joint2	Cross-condition	Joint1	Joint2
tw_2	LS-quadratic	0.86	0.91	1	0.81	0.85	0.92
tw_3	SMO-mlp	0.87	1	0.92	0.84	0.92	0.92
$tw_{1,2}$	SMO-quadratic	0.85	0.92	1	0.91	0.92	1
$tw_{2,3}$	LS-quadratic	0.91	1	1	0.88	1	0.75
$tw_{2,3}$	QP-quadratic	0.90	0.92	0.92	0.81	0.92	0.86
$tw_{2,3}$	SMO-mlp	0.95	0.86	0.92	0.85	1	0.91
$tw_{3,4}$	QP-linear	0.88	0.92	1	0.91	0.92	0.86
$tw_{3,4}$	SMO-quadratic	0.87	1	0.9	0.84	1	0.79
$tw_{3,4}$	LS-quadratic	0.86	0.86	1	0.81	1	1
$tw_{3,4}$	QP-quadratic	0.88	1	1	0.91	0.92	0.86

successful in detecting significant affiliations between features and class labels while avoiding overfitting. Furthermore, the high specificity and sensitivity values, as illustrated in Table 2, further corroborate the validity of the classifiers employed. The small numbers of both false positives and false negatives support the notion that there was no bias in favor of one class over the other during the classification of the actors' responses.

Concerning specific electrodes and features, no overall clear trend could be discerned. Nevertheless, as depicted in Table 3, for the 2 cases that *Joint1* accuracy reached 1 and *task-specific average* was higher than 0.9, the features selected to be added for *Joint1* presented a central/centro-parietal majority (9 of the 13 selected features). Moreover, for the 2 cases that *Joint2* accuracy reached 1 and task-specific average surpassed 0.9, the features selected for *Joint2* presented a parietal/centro-parietal majority (7 of the 8 selected features). Interestingly, in the above cases, a differentiation between the two condition-specific subsets was detected. In more detail, the condition-specific features added for *Joint1* condition were different from those added for *Joint2* condition, starting from the same SFFS cross-condition set, with an exception of one case, namely, feature MinT, for electrode Pz and tw_1 . The feature distributions for the two cases corresponding to the best classification results, i.e., cases where *Joint1* or *Joint2* accuracy was 1 and task-specific average had its highest value 0.96, are presented in Fig. 8.

4 Discussion

In this study, we performed cross-condition and within-condition classification on error-processing ERP signals in an auditory task with two levels of complexity. The presented framework was capable of selecting ERP characteristic features both common to the two conditions and separately for each condition, leading to successful discrimination between correct or incorrect

responses. In fact, although the waveforms of correct and incorrect responses – when averaged across all subjects and conditions – did not present a clearly distinguishable error-related differentiation (Fig. 2B) (as also indicated by previous research of our group on these data [38]), the high classification accuracy reached for cross-condition and within-condition classification corroborates our initial hypothesis that machine learning methods can successfully detect hidden patterns in ErrP features. Hence, incorrect decisions can be identified irrespective of the task difficulty, while additional ErrP characteristics that improve classification for each difficulty level can be extracted. Among the SVM models adopted in the present study, quadratic kernels presented the highest performance. Interestingly, rbf kernels failed to meet the performance evaluation criterion, suggesting that although the main advantage of SVM classifier is that – paired with the kernel trick – it can efficiently classify non-linear data, the fact that linear kernels present higher performance might indicate the linear nature of the features extracted [47]. This may well be the case, as other EEG classification studies also display better classification accuracy utilizing SVM kernels other than rbf [48–50]. For further validation, we also implemented the k-nearest neighbor (k-NN) and the linear discriminant analysis (LDA) classification techniques, using the methodology exposed above (see [Supplementary materials](#)). The performance of the k-NN and LDA classifiers was overall inferior to the SVM-based machine learning approach, although LDA reached acceptable performance levels, adding to the indications for the efficiency of employing linear modeling.

The response-related signals analyzed in our work are elicited after hearing FBT, i.e., the first feedback tone provided to the subjects. Therefore, the generation of an FRN-like signal might have been expected. Of note is that the overwhelming majority of error-related studies employ pre-defined time windows to detect and analyze error-related components [5, 19]. However, it should be taken into consideration that because of the nature of FBT, as

Table 3 Features selected

Time window	Classifier	Cross-condition	Features added for Joint1	Features added for Joint2
$tw_{2,3}$	LS-quadratic	tw_2 : FC2_MinA, Cz_MinT, Cz_MaxA, CP1_MaxT, CP2_MinA tw_3 : Fz_MinT, Cz_MinT, Cz_AUC, CP1_MinA	tw_2 : FC1_MinA, FC2_MaxT, P1_AUC, CP2_MinT, CP2_MaxA, Pz_MinT, Pz_MaxA tw_3 : FC2_MinA, Cz_MaxT, CP2_MaxA	tw_2 : FC1_MinT, FC1_MaxA, FC2_MinT, Cz_MaxT, CP1_MinA tw_3 : FC1_AUC, CP2_MinA, CP2_MaxT
$tw_{3,4}$	SMO-quadratic	tw_3 : Fz_MinT, FC2_MinA, CP1_MaxA, CP1_AUC, Pz_MinA, Pz_AUC	tw_3 : FC2_AUC tw_4 : CP2_MaxA, Pz_MinT	tw_3 : CP2_MaxT
$tw_{1,2}$	SMO-quadratic	tw_1 : FC1_MaxA, FC2_MinT, CP1_MinT tw_2 : Fz_MaxA, Fz_MaxT, Cz_MaxA, CP1_MaxT	tw_2 : Pz_MinT	tw_1 : FC2_MinA, CP1_AUC, CP2_AUC, Pz_MaxA tw_2 : Pz_MinT
$tw_{3,4}$	LS-quadratic	tw_3 : CP1_AUC tw_4 : FC2_MaxA	tw_3 : FC1_MaxT tw_4 : CP2_MinA, Pz_MinA, Pz_MinT	tw_3 : CP1_MinA, CP2_MinA tw_4 : CP1_MaxT

The first two cases correspond to the cases in Table 1 where *Joint1* classification accuracy reached 1, while the last two cases correspond to the cases in Table 1 where *Joint2* classification accuracy reached 1. Each feature is denoted including the electrode from which the feature was extracted, followed by the feature itself, i.e., “electrode feature”

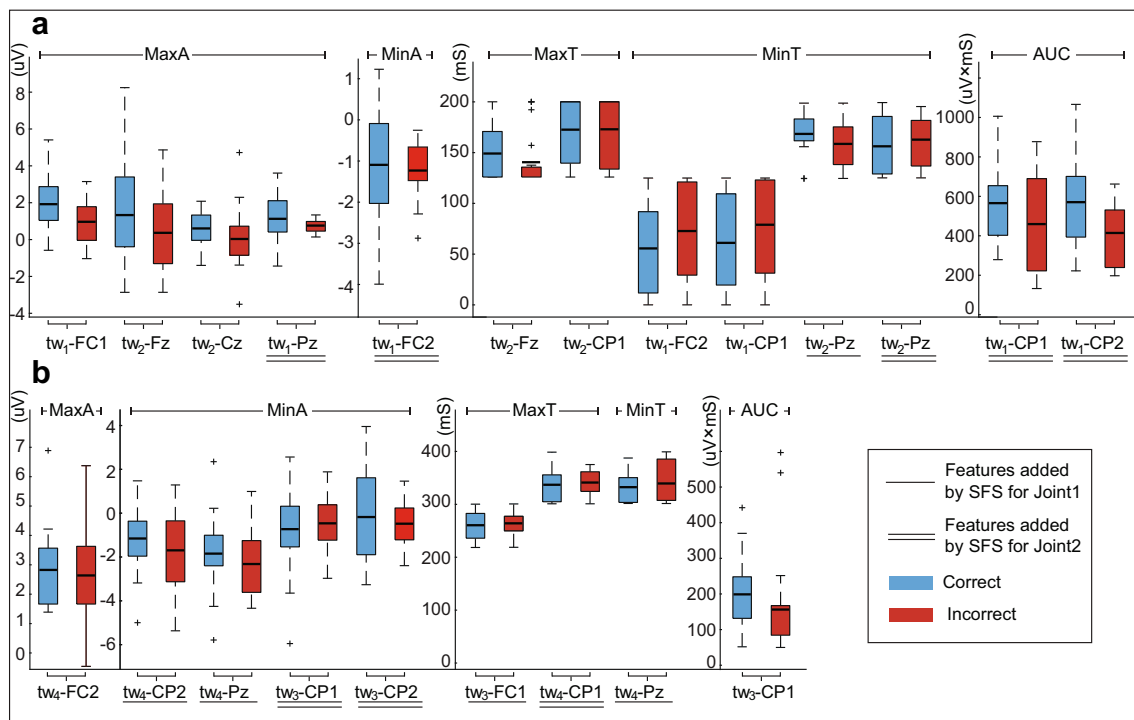


Fig. 8 Feature mean values and distribution for cases (a) $tw_{1,2}$, method, SMO quadratic, and (b) $tw_{3,4}$, method, LS quadratic. In each box, the central horizontal line indicates the mean value, while the edges of the

box indicate the 25th and 75th percentile. The whiskers extend to the most extreme data points and outliers are marked with the “+” symbol

explained in the methodology section, as well as the fact that amplitude and latency variations of ErrPs appear to depend on individual subject differences and task-condition manipulations [28, 51, 52], the morphology and duration of the error-related ERP signals could not be ascertained beforehand. Therefore, a series of consecutive time windows were investigated, into which features were computed, so as not to preclude latencies that could provide useful information. Results indicate that useful information can mainly be extracted from combinations of adjacent time windows $tw_{1,2}$ (0–220 ms), $tw_{2,3}$ (125–300 ms), and $tw_{3,4}$ (220–400 ms), instead of the short-duration single time windows. This can be related to the fact that the ErrPs corresponding to feedback tone processing can have error-related features extending in time windows of over 200 ms [53]. In addition, it could be inferred that, since the ErrPs tend to be distorted or masked by other components due to task complexity [54], the combination of time windows could provide a suitable approach for incorporating additional error-related components to the classification schemes. On the other hand, using an overly extended time window, i.e., tw_5 lasting from 0 to 600 ms, might confound the FS algorithms, as the large ERP peaks after 400 ms appearing in auditory tasks might reflect information unrelated to error processing and thus decrease the number of useful features [55]. In this context, it should be kept in mind that feedback-related ErrPs may be confounded and not be apparent due to variability of feedback valence and experimental conditions [56–58].

In addition, reward expectancy and reinforced learning effects modulate the characteristics of feedback-

related ErrPs, even in correct trials [59, 60]. The ERPs investigated in the current study originated from epochs where the FBT provided indirect information for the response of the actor. Therefore, the ERPs analyzed might not provide as clear error-related features as those that would have been extracted from ERPs recorded after the presentation of a sole feedback tone providing unambiguous information on the correctness of the participants’ actions.

Considering the features selected for the two conditions, in the cases where accuracy of *Joint1* or *Joint2* reached 1, the features selected by SFS for *Joint1* condition were different from the features selected by SFS for *Joint2* condition, starting from the same SFFS-selected set (see Table 3). Therefore, the feature sets that provided the best classification between ErrPs corresponding to correct and incorrect responses, although initiating from the same SFFS, when subsequently tailored to each condition separately resulted in sets differing for the two conditions, notably for the cases that provide the best classification accuracy. This is in line with other cross-condition pattern recognition studies presenting differentiations on ErrP classification performance related to condition manipulations [22, 30, 36]. In this regard, contrary to previously proposed methods that apply training and testing on different tasks [21–23], thus taking into account for training condition-salient features that subsequently result in impaired cross-condition performance, our framework succeeds in disentangling cross-condition and condition-specific

classification by selecting both common and individual condition features for the overall classification processes.

Additionally, the fact that the performance evaluation criterion and single-condition accuracy equal to 1 were met in several cases indicates that the procedures used in FS grant flexibility to the method, consequently providing high-accuracy results in classifying ErrP features corresponding to correct and incorrect responses from each condition. On the other hand, since FS took into account the mean value of not only cross-condition but also condition-specific classification accuracy, it might be expected that the feature set provided by SFFS, when only cross-condition classification was evaluated, would perform better for condition-independent classification, although bias might be introduced toward *Joint1* or *Joint2* classification.

Some considerations need to be taken into account when interpreting the results of the current study. In order to alleviate the effects of unbalanced conditions on SVM algorithms [61] and to elucidate generalizability in the evaluation of the error recognition, the average of each participant's ERP signals was employed for classification purposes, thus leading to a small number of instances to be classified.

Of note is that we only employed features that derive from morphological apparent signal characteristics (amplitude, latency, etc.), since our goal was to perform condition-independent and condition-dependent classification using rather simple ERP-based characteristics. Investigating more complex features might improve results.

Although the features selected for each method contribute to the classification accuracy, the degree to which they relate to the underlying condition-specific processes should be viewed with skepticism [62]. The main concern is the lack of consistency in the features selected, since for each method a different set of features was selected (see Table 3). In this context, the features that improve performance may not directly relate to the underlying neuronal processes and could have been chosen also because they allowed for reduction of noise unrelated to neuronal processing. Nevertheless, the existence of both condition-independent and condition-specific salient subsets of ErrP-based features might have the potential to successfully discriminate between ErrPs corresponding to correct and incorrect responses and provide indications for error-processing mechanisms adjusted to task difficulty [63, 64]. Taking the above into consideration, it can be conjectured that ErrPs associated with brain error-monitoring processes might reflect both elements of a common underlying error-detection cognitive mechanism and modifications of that mechanism, depending on the task complexity level. Toward this direction, we intent to extend this study in future work, in order to investigate the underlying brain mechanisms related to a universal error-processing mechanism irrespective of task complexity and elucidate neural substrates that regulate global and condition-specific error responses.

5 Conclusion

The cognitive error-related processing is deemed highly significant in the human behavior adaptation as well as in clinical research applications. However, even though ErrPs are stimulus-locked, they display large variations in signal characteristics as a result of different cognitive tasks and experimental conditions. As such, cross-condition error prediction based on ERP attributes remains a challenge. In this paper, we presented a framework for condition-specific and condition-independent classification of ERP-based features of an auditory identification task under two difficulty levels. Our analysis succeeded in providing a small number of feature subsets with high accuracy by utilizing a feature selection (SFFS-SFS) framework for handling both cross- and individual-condition error-processing variations, depending on the task complexity level. Results seem to support the initial hypothesis that machine learning algorithms employing a small number of ErrPs-based features have the potential to model hidden patterns and successfully discriminate between correct and incorrect responses in multiple conditions while additionally provide indications that the combinations of adjacent time windows can help incorporate ErrP components affected by complexity modifications.

Funding information This research is co-financed by Greece and the European Union (European Social Fund-ESF) through the Operational Programme «Human Resources Development, Education and Lifelong Learning» in the context of the project “Strengthening Human Resources Research Potential via Doctorate Research” (MIS-5000432), implemented by the State Scholarships Foundation (IKY).

Compliance with ethical standards

Conflict of interest The authors declared that they have no conflict of interests..

References

1. Luck SJ, Kappenman ES (2011) The Oxford handbook of event-related potential components. Oxford University Press
2. Wessel JR (2012) Error awareness and the error-related negativity: evaluating the first decade of evidence. *Front Hum Neurosci* 6:88. <https://doi.org/10.3389/fnhum.2012.00088>
3. Hewig J, Coles MGH, Trippe RH, Hecht H, Miltner WH (2011) Dissociation of Pe and ERN/ne in the conscious recognition of an error. *Psychophysiology* 48:1390–1396. <https://doi.org/10.1111/j.1469-8986.2011.01209.x>
4. Potts GF, Martin LE, Kamp S-M, Donchin E (2011) Neural response to action and reward prediction errors: comparing the error-related negativity to behavioral errors and the feedback-related negativity to reward prediction violations. *Psychophysiology* 48:218–228. <https://doi.org/10.1111/j.1469-8986.2010.01049.x>
5. Hauser TU, Iannaccone R, Stämpfli P, Drechsler R, Brandeis D, Walitza S, Brem S (2014) The feedback-related negativity (FRN) revisited: new insights into the localization, meaning and network

- organization. *NeuroImage* 84:159–168. <https://doi.org/10.1016/j.neuroimage.2013.08.028>
6. Vidal F, Hasbroucq T, Grapperon J, Bonnet M (2000) Is the ‘error negativity’ specific to errors? *Biol Psychol* 51:109–128. [https://doi.org/10.1016/S0301-0511\(99\)00032-0](https://doi.org/10.1016/S0301-0511(99)00032-0)
 7. Simons RF (2010) The way of our errors: theme and variations. *Psychophysiology* 47:1–14. <https://doi.org/10.1111/j.1469-8986.2009.00929.x>
 8. Keil J, Weisz N, Paul-Jordanov I, Wienbruch C (2010) Localization of the magnetic equivalent of the ERN and induced oscillatory brain activity. *NeuroImage* 51:404–411. <https://doi.org/10.1016/j.neuroimage.2010.02.003>
 9. Steele VR, Anderson NE, Claus ED, Bernat EM, Rao V, Assaf M, Pearson GD, Calhoun VD, Kiehl KA (2016) Neuroimaging measures of error-processing: extracting reliable signals from event-related potentials and functional magnetic resonance imaging. *Neuroimage* 132:247–260. <https://doi.org/10.1016/j.neuroimage.2016.02.046>
 10. Becker MPI, Nitsch AM, Miltner WHR, Straube T (2014) A single-trial estimation of the feedback-related negativity and its relation to BOLD responses in a time-estimation task. *J Neurosci* 34:3005–3012. <https://doi.org/10.1523/JNEUROSCI.3684-13.2014>
 11. Cohen MX (2011) Error-related medial frontal theta activity predicts cingulate-related structural connectivity. *NeuroImage* 55:1373–1383. <https://doi.org/10.1016/j.neuroimage.2010.12.072>
 12. Ullsperger M, Harsay HA, Wessel JR, Ridderinkhof KR (2010) Conscious perception of errors and its relation to the anterior insula. *Brain Struct Funct* 214:629–643. <https://doi.org/10.1007/s00429-010-0261-1>
 13. Iannaccone R, Hauser TU, Staempfli P, Walitza S, Brandeis D, Brem S (2015) Conflict monitoring and error processing: new insights from simultaneous EEG–fMRI. *NeuroImage* 105:395–407. <https://doi.org/10.1016/j.neuroimage.2014.10.028>
 14. Roger C, Bénar CG, Vidal F, Hasbroucq T, Burle B (2010) Rostral cingulate zone and correct response monitoring: ICA and source localization evidences for the unicity of correct- and error-negativities. *Neuroimage* 51:391–403. <https://doi.org/10.1016/j.neuroimage.2010.02.005>
 15. Meckler C, Allain S, Carbone L et al (2011) Executive control and response expectancy: a Laplacian ERP study. *Psychophysiology* 48:303–311. <https://doi.org/10.1111/j.1469-8986.2010.01077.x>
 16. Chavarriga R, del Millán JR (2010) Learning from EEG error-related potentials in noninvasive brain-computer interfaces. *IEEE Trans Neural Syst Rehab Eng* 18:381–388. <https://doi.org/10.1109/TNSRE.2010.2053387>
 17. Kim SK, Kirchner EA (2013) Classifier transferability in the detection of error related potentials from observation to interaction. In: 2013 IEEE international conference on systems, man, and cybernetics. Pp 3360–3365
 18. Zhang H, Chavarriga R, Khaliliardali Z, Gheorghe L, Iturrate I, Millán Jd (2015) EEG-based decoding of error-related brain activity in a real-world driving task. *J Neural Eng* 12:066028. <https://doi.org/10.1088/1741-2560/12/6/066028>
 19. Chavarriga R, Sobolewski A, Millán JDR (2014) Errare machinale Est: the use of error-related potentials in brain-machine interfaces. *Front Neurosci* 8:208. <https://doi.org/10.3389/fnins.2014.00208>
 20. Ventouras EM, Asvestas P, Karanasiou I, Matsopoulos GK (2011) Classification of error-related negativity (ERN) and positivity (Pe) potentials using kNN and support vector machines. *Comput Biol Med* 41:98–109. <https://doi.org/10.1016/j.combiomed.2010.12.004>
 21. Plewan T, Wascher E, Falkenstein M, Hoffmann S (2016) Classifying response correctness across different task sets: a machine learning approach. *PLoS One* 11:e0152864. <https://doi.org/10.1371/journal.pone.0152864>
 22. Iturrate I, Montesano L, Minguez J (2013) Task-dependent signal variations in EEG error-related potentials for brain–computer interfaces. *J Neural Eng* 10:026024. <https://doi.org/10.1088/1741-2560/10/2/026024>
 23. López-Larraz E, Creatura M, Iturrate I, et al (2011) EEG single-trial classification of visual, auditory and vibratory feedback potentials in brain-computer interfaces. In: 2011 annual international conference of the IEEE engineering in medicine and biology society. Pp 4231–4234
 24. Omedes J, Iturrate I, Montesano L, Minguez J (2013) Using frequency-domain features for the generalization of EEG error-related potentials among different tasks. In: 2013 35th annual international conference of the IEEE engineering in medicine and biology society (EMBC). Pp 5263–5266
 25. Balconi M, Crivelli D (2010) FRN and P300 ERP effect modulation in response to feedback sensitivity: the contribution of punishment-reward system (BIS/BAS) and behaviour identification of action. *Neurosci Res* 66:162–172. <https://doi.org/10.1016/j.neures.2009.10.011>
 26. Van den Berg I, Franken IHA, Muris P (2011) Individual differences in sensitivity to reward. *J Psychophysiol* 25:81–86. <https://doi.org/10.1027/0269-8803/a000032>
 27. Weinberg A, Dieterich R, Riesel A (2015) Error-related brain activity in the age of RDoC: a review of the literature. *Int J Psychophysiol* 98:276–299. <https://doi.org/10.1016/j.ijpsycho.2015.02.029>
 28. Iturrate I, Chavarriga R, Montesano L, Minguez J, Millán J (2014) Latency correction of event-related potentials between different experimental protocols. *J Neural Eng* 11:036005. <https://doi.org/10.1088/1741-2560/11/3/036005>
 29. Boldt A, Yeung N (2015) Shared neural markers of decision confidence and error detection. *J Neurosci* 35:3478–3484. <https://doi.org/10.1523/JNEUROSCI.0797-14.2015>
 30. Spüler M, Niethammer C (2015) Error-related potentials during continuous feedback: using EEG to detect errors of different type and severity. *Front Hum Neurosci* 9:55. <https://doi.org/10.3389/fnhum.2015.00155>
 31. Yousefi R, Sereshkeh AR, Chau T (2019) Online detection of error-related potentials in multi-class cognitive task-based BCIs. *Brain-Computer Interfaces* 6:1–12. <https://doi.org/10.1080/2326263X.2019.1614770>
 32. Luo T, Fan Y, Lv J, Zhou C (2018) Deep reinforcement learning from error-related potentials via an EEG-based brain-computer interface. In: 2018 IEEE international conference on bioinformatics and biomedicine (BIBM). Pp 697–701
 33. Hoffmann S, Falkenstein M (2010) Independent component analysis of erroneous and correct responses suggests online response control. *Hum Brain Mapp* 31:1305–1315. <https://doi.org/10.1002/hbm.20937>
 34. Kaczurkin AN (2013) The effect of manipulating task difficulty on error-related negativity in individuals with obsessive-compulsive symptoms. *Biol Psychol* 93:122–131. <https://doi.org/10.1016/j.biopsycho.2013.01.001>
 35. Kim KH, Kim JH, Yoon J, Jung K-Y (2008) Influence of task difficulty on the features of event-related potential during visual oddball task. *Neurosci Lett* 445:179–183. <https://doi.org/10.1016/j.neulet.2008.09.004>
 36. Endrass T, Klawohn J, Gruetzmann R et al (2012) Response-related negativities following correct and incorrect responses: evidence from a temporospatial principal component analysis. *Psychophysiology* 49:733–743. <https://doi.org/10.1111/j.1469-8986.2012.01365.x>
 37. Van der Borgh L, Houtman F, Burle B, Notebaert W (2016) Distinguishing the influence of task difficulty on error-related ERPs using surface Laplacian transformation. *Biol Psychol* 115:78–85. <https://doi.org/10.1016/j.biopsycho.2016.01.013>

38. Karanasiou IS, Papageorgiou C, Tsianaka EI, Matsopoulos GK, Ventouras EM, Uzunoglu NK (2009) Behavioral and brain pattern differences between acting and observing in an auditory task. *Behav Brain Funct* 5:5. <https://doi.org/10.1186/1744-9081-5-5>
39. Moore BC, Glasberg BR (1983) Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. *J Acoust Soc Am* 74:750–753
40. Ferrez PW, del Millan JR (2008) Error-related EEG potentials generated during simulated brain-computer interaction. *IEEE Trans Biomed Eng* 55:923–929. <https://doi.org/10.1109/TBME.2007.908083>
41. Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20:273–297. <https://doi.org/10.1023/A:1022627411411>
42. Theodoridis S, Koutroumbas K (2008) Pattern recognition, fourth edition, 4th ed. Academic Press
43. Pudil P, Novovičová J, Kittler J (1994) Floating search methods in feature selection. *Pattern Recogn Lett* 15:1119–1125. [https://doi.org/10.1016/0167-8655\(94\)90127-9](https://doi.org/10.1016/0167-8655(94)90127-9)
44. Jain AK, Duin RPW, Mao J (2000) Statistical pattern recognition: a review. *IEEE Trans Pattern Anal Mach Intell* 22:4–37. <https://doi.org/10.1109/34.824819>
45. Wong T-T (2015) Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. *Pattern Recogn* 48:2839–2846. <https://doi.org/10.1016/j.patcog.2015.03.009>
46. Hammerla NY, Plötz T (2015) Let's (not) stick together: pairwise similarity biases cross-validation in activity recognition. In: Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing. ACM, New York, pp 1041–1051
47. Theodoridis S, Koutroumbas K (2008) Pattern recognition, 4th edn. Academic Press, Amsterdam
48. Singla R, Chambayil B, Khosla A, Santosh J (2011) Comparison of SVM and ANN for classification of eye events in EEG. *J Biomed Sci Eng* 04:62–69. <https://doi.org/10.4236/jbise.2011.41008>
49. Parvar H, Sculthorpe-Petley L, Satel J, Boshra R, D'Arcy RC, Trappenberg TP (2014) Detection of event-related potentials in individual subjects using support vector machines. *Brain Inform* 2:1–12. <https://doi.org/10.1007/s40708-014-0006-7>
50. Nicolaou N, Georgiou J (2012) Detection of epileptic electroencephalogram based on permutation entropy and support vector machines. *Expert Syst Appl* 39:202–209. <https://doi.org/10.1016/j.eswa.2011.07.008>
51. Hughes G, Yeung N (2011) Dissociable correlates of response conflict and error awareness in error-related brain activity. *Neuropsychologia* 49:405–415. <https://doi.org/10.1016/j.neuropsychologia.2010.11.036>
52. Grützmann R, Endrass T, Klawohn J, Kathmann N (2014) Response accuracy rating modulates ERN and Pe amplitudes. *Biol Psychol* 96:1–7. <https://doi.org/10.1016/j.biopsycho.2013.10.007>
53. Baker TE, Holroyd CB (2011) Dissociated roles of the anterior cingulate cortex in reward and conflict processing as revealed by the feedback error-related negativity and N200. *Biol Psychol* 87:25–34. <https://doi.org/10.1016/j.biopsycho.2011.01.010>
54. Gawłowska M, Domagalik A, Beldzik E, Marek T, Mojsa-Kaja J (2018) Dynamics of error-related activity in deterministic learning - an EEG and fMRI study. *Sci Rep* 8:14617. <https://doi.org/10.1038/s41598-018-32995-x>
55. Choudhury NA, Parascando JA, Benasich AA (2015) Effects of presentation rate and attention on auditory discrimination: a comparison of long-latency auditory evoked potentials in school-aged children and adults. *PLoS One* 10:e0138160. <https://doi.org/10.1371/journal.pone.0138160>
56. Ferdinand NK, Mecklinger A, Kray J, Gehring WJ (2012) The processing of unexpected positive response outcomes in the mediofrontal cortex. *J Neurosci* 32:12087–12092. <https://doi.org/10.1523/JNEUROSCI.1410-12.2012>
57. Kreussel L, Hewig J, Kretschmer N, Hecht H, Coles MG, Miltner WH (2012) The influence of the magnitude, probability, and valence of potential wins and losses on the amplitude of the feedback negativity. *Psychophysiology* 49:207–219. <https://doi.org/10.1111/j.1469-8986.2011.01291.x>
58. Opitz B, Ferdinand NK, Mecklinger A (2011) Timing Matters: The Impact of Immediate and Delayed Feedback on Artificial Language Learning. *Front Hum Neurosci* 5:8. <https://doi.org/10.3389/fnhum.2011.00008>
59. Krigolson OE, Hassall CD, Handy TC (2014) How we learn to make decisions: rapid propagation of reinforcement learning prediction errors in humans. *J Cogn Neurosci* 26:635–644. https://doi.org/10.1162/jocn_a_00509
60. Chase HW, Swainson R, Durham L, Benham L, Cools R (2011) Feedback-related negativity codes prediction error but not behavioral adjustment during probabilistic reversal learning. *J Cogn Neurosci* 23:936–946. <https://doi.org/10.1162/jocn.2010.21456>
61. Sun Y, Wong AKC, Kamel MS (2009) Classification of imbalanced data: a review. *Int J Pattern Recognit Artif Intell* 23:687–719. <https://doi.org/10.1142/S0218001409007326>
62. Haufe S, Meinecke F, Görgen K, Dähne S, Haynes JD, Blankertz B, Bießmann F (2014) On the interpretation of weight vectors of linear models in multivariate neuroimaging. *Neuroimage* 87:96–110. <https://doi.org/10.1016/j.neuroimage.2013.10.067>
63. van Driel J, Ridderinkhof KR, Cohen MX (2012) Not all errors are alike: theta and alpha EEG dynamics relate to differences in error-processing dynamics. *J Neurosci* 32:16795–16806. <https://doi.org/10.1523/JNEUROSCI.0802-12.2012>
64. Gentsch A, Ullsperger P, Ullsperger M (2009) Dissociable medial frontal negativities from a common monitoring system for self- and externally caused failure of goal achievement. *Neuroimage* 47:2023–2030. <https://doi.org/10.1016/j.neuroimage.2009.05.064>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Ioannis Kakkos received a B.S. degree in Physics (2013) and an MSc degree (2015) in Biomedical Engineering from the University of Patras. He worked as a research assistant with the Singapore Institute for Neurotechnology (SINAPSE) in several brain research projects. Currently he is pursuing a Ph.D. in Cognitive Engineering in the National Technical University of Athens. His research interests include the transmission of nerve stimuli, the study of cognitive systems and processes, and the development of medical treatment techniques for cortical signal processing and rehabilitation.



Errikos M. Ventouras got his diploma from the Department of Electrical and Computer Engineering of the National Technical University of Athens (NTUA) in 1989. He completed his doctoral thesis at the same Department in 1994. From 1994 till 1998, he worked as research lecturer in the University Research Institute of Computer and Communications of NTUA and in private sector R&D departments. From 1998 he is with the Department of Biomedical

Engineering of the University of West Attica, where he is currently a Professor. He has taught at graduate courses related to Biomedical Engineering at the University of West Attica, the University of Patras, and the University of Athens. His research interests focus on acquisition and diagnostic utilization of brain electrical signals, biosignal digital processing, cognitive neurodynamics, classification techniques, tomographic techniques for medical imaging, and electronic examination methods. He has 37 publications in refereed international journals and over 130 publications in book chapters and conference proceedings and abstracts.



Pantelis A. Asvestas got his diploma in engineering from the Department of Electrical and Computer Engineering of the National Technical University of Athens in 1996 and the title of Doctor in Engineering from the same Department in 2001, on medical image processing. From 2003 till today, he has been working as an electrical engineer, participating in R&D projects. He is an associate professor in the Department of Biomedical Engineering of the University of

West Attica, Greece. His research interests cover classification of biosignals and medical images using artificial intelligence, image processing, registration and fusion of medical data, 3D visualization of medical data, biometrics, and real-time video processing. He has 47 papers in international scientific journals and 74 scientific publications in book chapters, proceedings of international or national congresses, and national scientific magazines.



Irene S. Karanasiou was born in Athens, Greece. She received her diploma and her Ph.D. degree in Electrical and Computer Engineering from the National Technical University of Athens (NTUA), Athens, in 1999 and 2003, respectively. Since 1999, she has been a researcher with the Microwave and Fiber Optics Laboratory (MFOL), NTUA, and currently she is associate professor at the Hellenic Military University. She has authored or co-authored more than 150 papers

in refereed international journals and conference proceedings. She was the member of the organizing and technical committees of more than 20 international conferences including IEEE conferences. She has participated in more than 24 funded national and European research projects. Her current research interests include biomedical imaging techniques, medical informatics, bio-electromagnetism, and applications of microwaves in therapy and diagnosis. She is also working in the field of electroencephalography (EEG), evoked potentials (ERPs) and computational neuroscience, functional brain imaging (fMRI), and recently terahertz technology and imaging. Dr. Karanasiou is a founding member of the IEEE Engineering in Medicine and Biology Society (EMBS) Greek Chapter and member of the Technical Chamber of Greece. She was the recipient of the Thomaidio Foundation Award for her doctoral dissertation (2004) and three academic journal publications.



George K. Matsopoulos received his diploma in electrical engineering in 1988 from the National Technical University of Athens. He received his M.Sc. in 1989 from the Bioengineering Unit, University of Strathclyde, UK, with thesis title “Location of the femur in Above-Knee Amputee wearing Quadrilateral and Ischial Containment Socket using Ultrasound”. He received his Ph.D. in Bioengineering Unit in 1993 from the University of Strathclyde, UK, with thesis title

“Application of Morphology to Medical Image Processing”. He worked as a product specialist of Toshiba’s Medical Imaging Modalities at ERGO S.A. Company, one of the leading in medical equipment, for several years. He is currently a professor at the School of Electrical and Computer Engineering of the National Technical University of Athens. His research interests include signal processing, image processing, 2-D and 3-D registration and fusion, biosignal processing, optimization techniques, decision support systems, telemedical and computer vision applications, and hospital information systems. He has published over 300 papers in international scientific journals, chapters in books, and in international and national conference proceedings.