A study on the utility of temporal derivatives and unsupervised clustering in brain-computer interfaces

Dimitrios Andreou

A thesis submitted for the degree of Master of Philosophy (MPhil)

School of Computer Science and Electronic Engineering
University of Essex

April 19, 2017

Dedicated to my grandparents,
those currently fighting their age who are constantly reminding me of who I am,
the one I lost during my studies – may her cheerfulness always inspire me –
and the one I never met but feel I have always known.

Acknowledgements

I would like to thank my supervisor Riccardo Poli for all his work, the time he dedicated to me, his feedback and foresight. I would like to thank my colleagues Ana Matran Fernandez and Davide Valeriani whom I learned much from. Many thanks to my mother Sofia Geki, my father Alexandros Andreou, my brother Antonis Andreou and my beloved friends for their perseverance. Last but not least, I would like to extend my gratitude to my English teacher Salomi Vasileiadou whose help was invaluable.

Abstract

Brain-computer interfaces (BCIs) rely on accurate classification of event-related potentials (ERP), a task commonly delegated to a machine-learning algorithm, which investigates features derived from the voltages (V) recorded at different scalp locations with the electro-encephalogram (EEG). The performance of the machine-learning algorithm is an area that has captured the interest of the research community. Although major advancements have been made, BCIs suffer from uncertainties that arise from assumptions such as that participants are "focused", "still" and that no unpredictable events occurred during the recording, for example abrupt sounds or light changes.

From the range of possible uses of BCIs, one of the most challenging is its adaptation to everyday life situations. Addressing both participant and environmental related influences to the EEG could enable the usage of BCIs outside the confines of the laboratory. In addition, in order to create a BCI that can act as an "enhancement" for the able-bodied requires a way to identify recurrent events without prior knowledge, thus providing the user with a way to increment the "understanding" of his BCI. Moreover, information such as location, latency and shape of recurring events could provide solid grounds for future researchers to build upon.

In the thesis the above problem is challenged by investigating two main topics: assuming that the neuro-signals are additive (i.e. uncorrelated), (a) the usage of the first time derivative of V (dV) as feature regarding performance in classification of an ERP, and (b) unsupervised clustering of ERPs. Both investigations tackle the problem of mining properties of unknown neuro-signals. Theoretical investigations carried out on in each topic are performed using synthetic signals to assess the expected behaviour.

Using real data from a P300 BCI mouse, both topics were evaluated; the classification performance of dV was found to be significantly better than V while evaluating a baseline for comparison. Having such a positive outcome encouraged an attempt to create a single linkage unsupervised clustering method based on statistical significance. Without knowing if an ERP was generated or not, the developed clustering algorithm, based on dV, is shown to be accurate in identifying the shape of the underlying, "unknown" ERP.

For years researchers have been constructing experiments to uncover EEG events directly related to stimuli. An outcome of this research is that recurring EEG responses which might have been neglected, simply because they were not expected, are now identifiable.

Contents

List of Publications

Dimitrios Andreou and Riccardo Poli. Clustering simulated Event-Related Potentials based on similarity of centroids. In *Computer Science and Electronic Engineering Conference (CEEC), 2014 6th*, pages 144–147. IEEE, 2014

Dimitrios Andreou and Riccardo Poli. Comparing EEG, its time-derivative and their joint use as features in a BCI for 2-D pointer control. In *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference*, pages 5853–5856. IEEE, 2016

A Stoica, Ana Matran-Fernandez, Dimitrios Andreou, Riccardo Poli, Caterina Cinel, Y Iwashita, and C Padgett. Multi-brain fusion and applications to intelligence analysis. In *SPIE Defense, Security, and Sensing*, pages 87560N–87560N. International Society for Optics and Photonics, 2013

List of Figures

List of Tables

Chapter 1

Introduction

Being able to control machines without any musculoskeletal activity through brain-computer interfaces (BCI) can have multiple applications [4, 5]. Assisted living technologies based on this could provide people with severe neurological injuries with a means to a life unhindered by their condition [6]. The earliest BCI tackled communicating, giving the user spelling capabilities only by means of the electroencephalogram (EEG) [7]. Since the first appearance of the BCI spellers, researchers have extended the capabilities of these interfaces and tested various other applications such as robotic control and navigation [8, 9]. Using EEG and understanding the response to stimulation (event-related potentials, ERPs) on the basis of location and latency is required for the aforementioned BCIs to perform accurately. Either through transformations or selected properties, one can generate features that can be used to identify the above-mentioned responses. The Speller exploited the voltage features of the EEG (V), and later publications introduced their concepts based on these. BCIs have yet to be proven efficient in the environment outside the confines of the laboratory.

The current state of BCIs makes it very hard for them to be efficient in an every-day scenario when used by the able-bodied as an enhancement. Both classification performance and information transfer rates are being successfully improved by machine learning and a careful construction of stimulus presentation, but BCIs are still confined to recognise events solely based on *a priori* information. In the case of a BCI that aims to enhance the capabilities of an able-bodied user, EEG responses not only need to be accurately identified, but also recognised autonomously, thus providing a way for the user to evolve the BCI system through its usage. Evolving the usage of the BCI could possibly come in the form of a finer classification of a single response, or as an addition of more, previously unknown responses, that can be associated to more machine actions. Assuming that a user can monitor occurring events, if a response was generated due to an event, a combination of machine algorithm and features can accurately identify that response, but acquiring this *a priori* information is challenging.

Feedback-oriented BCIs aim at increasing the accuracy of classification by helping users identify times where they were correct or incorrect [10, 11]. Such BCI systems have been employed in rehabilitation. Since BCIs are restricted to the usage of known information/characteristics, they cannot encompass new information, such as stimuli that have not previously been defined but are presented to

the user through the environment. The strict pre-defined setups commonly used during laboratory experiments, which aim at having explainable results, mitigate the problems of unpredictability and evaluation of performance. That is, in a laboratory setting one can argue that the EEG responses are a direct result of the experiment. When the BCI is therefore taken out of the laboratory, it is uncertain whether the EEG responses are relative to the experiment, the environment, or both. Without an unsupervised process of identifying responses, there is no way of distinguishing between environmental and experimental stimuli with certainty. Evaluating the performance of the BCI outside the laboratory, then, is currently biased since the user now feeds on more information than that solely produced by the experiment specifications. Thus, when evaluating the performance of the BCI one can only accept this bias with no means of handling it.

In this thesis we evaluate the relative benefits of using the first order temporal derivatives of V (dV) as input to the BCI. The usage of dVs in this manner is beneficial, at least for stimuli that produce responses with large deviations in terms of V. ERPs produced by such stimuli are bound to produce ripples which correspond to large changes in the potentials of the electric field. The contribution in terms of classification of dV-based features in BCIs has not yet been thoroughly investigated and as such it is of interest to evaluate the relative benefits of these features.

In general, derivatives have been used before in EEG analysis [12, 13, 14], mostly as a tool for identifying rapid inflexions commonly produced during epileptic seizures. Differentiation as a measure of rapid change has also been used in detecting haemoglobin changes in order to enhance Steady State Visually Evoked Potential (SSVEP) classification [15].

In BCIs the usage of derivatives in literature is limited. Least square estimates of derivatives were used as features transformed via Principal Component Analysis (PCA) in a BCI matrix speller [16, 17]. There they were found to provide benefits both in terms of training times and reduced classification errors over amplitude-based features. However, no statistical test was performed to assess the significance of this finding. In addition, combining V and dV features as in the aforementioned publications produces no information on the performance of dV-based features. Thus, the question of choice between V and dV features remains unanswered.

From another point of view, a relation exists between differentiation and wavelet analysis: reverse biorthogonal wavelets, e.g. rbio3.1, can be used to extract regularised estimates of the first order derivative. Wavelets have been used in BCI research in a number of ways, e.g. to create synchronous [18] and asynchronous [19] BCIs. Also, the use of the continuous wavelet transform is a widely accepted method for generating features which can later be used to classify ERPs [20, 21, 22, 23, 24]. To the best of our knowledge, however, the interpretation of wavelets as regularised derivatives has not been considered in BCIs. It may be possible that an automatic production of wavelets used in feature generation had derivative-like results.

In this thesis, I aim to ascertain if dV features hold information beneficial to BCIs. In addition to a baseline examination of dV as a feature, I will also be examining the case of extracting *unknowingly* produced event-related potentials. To show the viability of this premise, a single-linkage hierarchical clustering method based on the Kruskal-Wallis test was created, while knowing a priori only of the

timing of an event, not their respective labels.

To begin, in the literature review (chapter 2), findings using information from V feature vectors, information regarding BCIs and processing of the EEG is presented, concluding with information regarding the data sets used in the analyses carried out in the thesis. The chapter aims to show the current focus of the research in order for the reader to appreciate the novelty of the enhancement-BCIs concept, while also providing a general understanding of the area. In chapter 3, first an investigation of the feature bases (V,dV) is carried out (latency, location and bandwidth) in terms of classification performance measured by the area under the receiver operating characteristics curve (AUC). Secondly, statistical evidence of the superiority of dV over V features is given. In chapter 4, an investigation of unsupervised clustering methods and criteria is carried out followed by an evaluation of an unsupervised clustering algorithm applied to both V and dV. Finally, in chapter 5 some conclusions are drawn and ideas regarding further investigation of dV features and clustering are presented.

Chapter 2

Literature Review

2.1 Information Transfer and Processing in the Brain

Information processing in the brain can be glimpsed at by capturing the changes in the electrical field using the EEG at the scalp of the participant. Potential differences between electrodes unveil the collective activity of similarly-oriented neurons, including background activity at different locations of the scalp. Neurons not only transmit information involuntarily as a result of stimuli (event-related potentials, ERPs)[25], but also on demand, e.g. thinking of moving the body's limbs (motor imagery, MI)[26]. ERPs and MI are examples of recognisable variability captured by the EEG; steady-state evoked potentials (SSEP) [27] and major rhythms [28] such as the α rhythm are also contained in EEG recordings. Each of the above reflects aspects of the functioning of the brain and although they are usually separately characterised in literature, in reality they may occur concurrently.

Major rhythms found in the EEG are usually denoted by α, β, γ, δ and θ. These EEG fluctuations lie approximately within the following frequency ranges: α between $8-12$ Hz, β between $12-30$ Hz, γ between >30 Hz, δ between $0.5-4$ Hz and θ between $4-8$ Hz [28]. Such oscillations are common between people, which suggests that the brain makes use of oscillation generators in order to carry out various operations [29]. Pathological conditions have been found to modulate these rhythms. For instance, autism has been found to influence the α rhythm; autistic subjects are able to perceive a larger amount of information from a stimulus when compared to healthy subjects [30, 31]. This in turn suggests that the α rhythm could be related to inhibiting brain activity.

In an attempt to explain these variations in the EEG, research indicates there is a 'top-down'/'bottom-up' way of communication between regions in the brain that acts as a means of synchronisation between neuron complexes [32, 33]. The highly inter-connected neurons of the brain make use of tolerances (at the synapse) for the propagation of information when communicating with each other. The post-synaptic neuron's membrane potential, when surpassed by the pre-synaptic neuron's potential level, allows for information transferred. These membrane potentials as described in literature are not constant in all contexts and environments, but rather vary due to several factors, such as those described in [34] for decision-making tasks.

Motivating the usage of machine learning in the topic of neuro-signal recognition is the high complexity of the recordings. The brain is a part of the human living organism and operates constantly throughout life. Chemicals affect the EEG readings, e.g. smoking [35]. The total resistance due to the material, composition and structure of the hair, skin, skull [36, 37, 38], meninges, brain matter and brain lesions [39, 40] also affect the EEG. Additionally, the functionality of internal organs introduces a variable influence in the EEG. Variable abnormalities in the EEG such as those created by cardiac and respiratory cycles affect the measurements not only due to the brain functioning to maintain their operation but also by their effect on the movement of the brain: with both respiratory and cardiac rhythms, the brain moves as a result of changes in blood vessel pressure. Modulations in respiratory cycles introduce EEG fluctuations [41] while modulations in cardiac rhythms can be correlated with conscious arousal [42].

2.2 Brain Computer Interfaces

Brain Computer Interfaces (BCIs)[7] utilise signal acquisition techniques to capture brain activity (such as EEG) and signal processing to trigger actions from human to machine. BCIs rely on identifying an EEG response with known psycho-physiological properties, i.e. the latency, location and bandwidth of the response. The connectivity characteristics of recognised EEG responses also contribute to BCIs. In MI BCIs, for instance, responses could occur in the right or left hemisphere depending on the laterality of the body part imagined. Hybrid BCI attempts [43], i.e. BCIs that combine the EEG with multiple data acquisition techniques for their operation, have been successful to some extent. BCIs are currently implemented primarily as an assistive technology to tackle a range of impairment issues, e.g. trauma rehabilitation and the robotic hand application [44].

The authors of the above-mentioned publication suggested that the challenge is to dynamically estimate and represent the user's intentions relative to a changing environment. A changing environment is challenging as random stimuli are presented to the user at unknown times. A constant environment where stimuli are timed and controlled poses the challenge of interpreting the meaning of the responses, whereas in the open environment this challenge is shadowed by increased difficulty in identifying the responses in the EEG.

Some reasons [44] for restrictions in the application and effectiveness of most BCIs are low information transfer rates between the operator and the machine, the inefficiency of pattern recognition capabilities under noisy conditions, the inconvenience of the required equipment and the missing knowledge of unexplained variation in signals.

BCIs mainly follow a common pattern of operation. The way in which the machine will be interacting with the participant is commonly referred to as the paradigm. A paradigm contains a set of rules and stimulation characteristics for the participant. For instance: "You will first hear a beep which will indicate that the machine has started working (focus request). As soon as the machine starts, mentally visualise moving one of your limbs (task). Keep the movement and the choice of limb constant, try not to move, blink or swallow while you visualise (reduce EEG complexity)." In the above example the user was

asked to mentally visualise a movement. Cues such as "(focus request)"and "(task)" are commonly used to inform the participant when to operate under the paradigm. There are many different types of BCIs, but all fall into two categories, synchronous and asynchronous [1]. Synchronous BCIs make use of cues not only while training the machine to identify a specific response but also while the BCI is in use. Asynchronous BCIs use cues only during the training of the machine; later the BCI is used without cues. It is common to instruct participants to keep their movements to a minimum since muscle movement produces artefacts that hinder the machine learning process.

In order for the machine to be able to efficiently identify an EEG component and return the classification, a series of steps is carried out - the BCI pipeline. The pipeline differs between training the machine and using the BCI. Varying EEG responses are distinguishable in different ways depending on the physiological characteristics of their generation, e.g. oscillatory (MI) and synchronised firing (ERP). Hence, a BCI researcher has to adapt EEG processing to fit the nature of the response. That being said, there are some common steps needed to access descriptive information of the chosen ERP response. The common BCI pipeline starts by splitting the EEG recording into segments according to the cues used in the paradigm; the resulting segments are called epochs. Since electrical activity can only be viewed through potential difference the next step is to reference the epochs. Referencing usually makes use of the average potential from the two earlobe electrodes. Although this referencing procedure is commonly used, a different selection of reference electrodes can also be used, e.g. referencing to the mean of all electrodes. With the EEG recording epoched and referenced the final step is filtering, which aims to remove noise and select bands where the EEG response is most dominant. Once the epochs have been filtered, the next portion of the pipeline usually considers artefact removal. Artefacts are generated for a number of reasons stemming from the environment - screen flickering for instance - or from involuntary muscle movement - swallowing, eye blinks and movement, cardiac rhythms, and so on. Their impact on the EEG is often visible even without processing of the signals and their effect on machine learning and classification is enormous. There are two main ways to handle artefacts in BCIs [45]: one can either identify epochs where artefacts have distorted the signals and reject them, or attempt to restore responses in the epochs by removing the effects of the artefacts.

The researcher now has epochs which are referenced, filtered and less contaminated by artefacts. This is the stage at which BCI researchers usually extract features from the epochs. Feature extraction has been a step at the core of BCI research since its very beginning [46, 47, 48, 49, 50]. Since this thesis also evaluates the usage of the derivative as feature, a separate section is devoted to feature selection 2.4. Once the features have been extracted from each epoch, the next step is usually referred to as the training of the BCI; this focuses on machine learning [51]. The use of machine learning is very common in BCIs, the main reason being its accuracy, despite the question of efficiency arising by the training required. There is no unique machine learning classification algorithm used in BCIs, but common preferences are evident throughout the literature. Linear classifiers such as Support vector machines (SVMs) and the linear discriminant (LDA) are amongst the most popular in BCIs. The machine learns the feature composition of the epochs – one representing a hand movement for instance – and is able to produce a

[1] In literature this categorisation can also be found as cue-based BCIs and non cue-based BCIs.

classification for unseen data thereafter.

2.3 Event Related Potentials

Neurons within the brain tend to synchronise their firing in response to a specific event; this produces measurable voltage variations on the scalp, which are called event related potentials (ERPs) [52]. Hence ERPs are time-locked to a sensory event, i.e. a stimulus. Mental processes are associated to such events and the EEG provides a means of capturing such ERPs. ERPs that appear approximately within 100 ms after stimulus presentation are termed "sensory" or "exogenous"; physical stimuli parameters have been found to vary in these early ERPs [53]. Later ERPs are thought of as the subject's evaluation of the stimulus. Such ERPs are commonly referred to as "cognitive" or "endogenous" [46].

Currently ERPs are characterised by their latency and amplitude, i.e. when they appear relative to the stimulus and the intensity of the reading. These are found to vary according to the psychological state of the subject and stimulus characteristics. They can be told apart by their names as follows: if the wave is positive then the prefix *P* is used followed by its approximate occurrence time after stimulus onset. In contrast, if the wave is negative, then the prefix *N* is used and is similarly followed by its occurrence time stamp. By using this convention, known waves can be distinguished, e.g. N200, P300. A second convention for naming concerns a wave's peaks; in addition to *P* or *N*, a number indicates the order of the peaks time-wise, e.g. a first positive would be P1. The components of the wave are usually denoted by a suffix letter; for instance for the auditory N200 wave (using the first convention), the first component is denoted by N2a (using the second convention, also found as mismatch negativity, MMN).

There is a vast amount of ERPs that could be considered for the purposes of the thesis. Nevertheless, to compare the derivative of the EEG as a feature, the counterpart of the comparison must be a feature with known capabilities, i.e. the amplitude. Since amplitude features are very commonly used in P300-based BCIs, selecting the P300 ERP would be suitable for the above-mentioned comparison.

2.3.1 The P300 Wave

As described in the previous section, the label "P300" refers to a positive inflection in the EEG which occurred about 300 milliseconds after stimulus presentation. As a neuro-signal, the P300 is the synchronised activation of neurons from a stimulated region of the cortex.

This cognitive wave appears when a mental task is associated with a stimulus. The P300 wave has been observed in contexts other than BCIs [54, 55, 56], but as an example, the BCI-"speller" [7] will be used. During this experiment, a participant with the EEG measuring their brain activity is positioned facing a screen that depicts an array of letters flashing in random order. Prior to the trial the participant is instructed to focus on a specific letter. If the column/row of the letter in focus flashes, the EEG recordings show a positive fluctuation about 300 ms after the stimulus (flashing) has occurred.

The amplitude of the P300 wave has been found to vary due to inter-stimulus interval, sequence length and target probability. More precisely, as described in [57], if stimuli are presented rapidly the amplitude of the P300 wave and its latency decrease. Conversely, as the rate of stimulus presentation

decreases, the amplitude of the P300 wave increases. The element of surprise also plays a role as the amplitude of the P300 wave is greater in situations were the stimulus appearance cannot be predicted from previous stimuli. [2]

The P300 has two components, P3a [59] and P3b, associated with short-term memory updates. This is the context update theory [60], which states that the P300 triggers the brain to update its context due to a change, signified by the stimulus. The discrimination between P3a and P3b is that the former is elicited by task-relevant stimuli, and the latter by a go-no go classification task [61]. Focusing on a particular letter in the speller paradigm, therefore, creates a P300 wave signifying the need to update the context due to the occurrence of a predefined requested change – the flash. If the subject was directed to keep a mental count of the number of flashes, then a P3a would be generated.

There have been considerable advancements in our understanding of the characteristics of the P300 wave, using EEG (high temporal resolution) and joint EEG and fMRI (high spatial resolution) experiments [57, 58, 62, 63, 64]. Joint EEG and fMRI analysis on P300 components has been conducted by [62], and they have shown a clearer representation of both the neuronal network communication and the occurrence of ERPs. A well aimed approach built on evidence from prior experiments revealed information on stimuli relevant to the user (self-relevant) in the context of the P300 being seen as an index of attentional resources [65]. The authors showed that when elicited by stimuli relevant to the user, such as a picture of a family member, the amplitude of the P300 wave was higher than that of the P300 wave elicited by non user-relevant stimuli – a picture of an unknown person, for instance. This indicated that the brain might use the wave to allocate resources such as memory or simply that the wave is a manifestation of the process.

Previous investigative work summarised by [60] has also hinted that the P300 originates from task definitions which involve the use of working memory. When the subject is instructed to watch a letter and count the number of flashes of that letter, it has been found that the P300 amplitude decreases when the primary task difficulty is increased. Also, ultradian rhythms have been found to contribute to P300 variation in latency and amplitude [66, 67]. These rhythms are recurrent cycles throughout a 24-hour circadian day consisting of a collection of biological processes such as testosterone secretion and bowel movement, which display oscillations of approximately one day.

In [68], the amplitude and latency of the P300 wave has been investigated in subjects with normal and obscured vision. Results suggested that the amplitudes for subjects whose vision was obscured were lower than for those with normal vision. Obscured vision is characterised by loss of information. According to the context update theory, the P300 would signify a smaller amount of resources needed to process the information received, when compared to normal vision subjects.

Experiments conducted on twin and family studies based on amplitude and latency of the peak of ERPs suggest that genetic similarity can affect ERP similarity [69]. Thus, structural/functional similarities of brain neuronal complexes between subjects are likely to become apparent in subjects with similar genetic makeup.

Another topic of research that has received attention mainly due to the difficulty it produces in P300

[2] Work on the P300 amplitude also highlights the subjective and objective probability effects [58].

measurements is eye-blinks [70]. Along with other ocular artefacts such as eye-moving muscle interference, eye-blinks during a P300 experiment can severely distort ERPs. Unfortunately the probability of blinking during an experiment is not small [71, 72]. Eye-blinks are also of general scientific interest as their occurrence can be intentional or unintentional [73] and their function is associated with clearing short-term memory.

Several studies tackle the interpretation of the P300 wave and its components, and the above is a very small sample of the literature available on the topic. As suggested initially, the importance of the P300 wave in this thesis is that it is hidden within the noise; 'within' here suggests that the signal to noise ratio approaches unity, i.e. the noise amplitude is equal to the signal amplitude. As discussed in later sections, this fact has resulted in the interpretation of ERPs (in an EEG experiment) being restricted to averages.

2.4 Features of Event Related Potentials

Attaining complete control of the machine without limb movement is the primary goal of BCIs. The enormous quantity of information and complexity in inter-communication of the brain has created a need for powerful, reliable feature extraction methods. Some which are considered in modern ERP feature extraction and selection literature are stated below.

The main stumbling blocks these methods tackle are the accurate classification of an ERP, their distinguishability and a reduction in dimensionality. Even current state-of-the-art classification methods cannot cope with the vast amount of information at hand. Hence, in part feature extraction encompasses areas such as feature selection [74]. The most common form of feature selection in literature is sub-sampling, widely used to narrow down the choice of data in order for classification to be optimised and avoid over-fitting when classification algorithm use is based on machine learning. Descriptive features, e.g. [75], are usually extracted depending on the nature of the response, i.e. requiring prior information to be known regarding the EEG response.

Feature extraction has resulted nowadays in a set of steps. Re-referencing is performed prior to filtering, followed by down-sampling, thus accounting for aliasing. Then a desired window is selected, usually based on a cue or label which represents events such as stimulus onset and response. This process is known as epoching (sometimes referred to as "windowing"). The resulting epochs are then further processed depending on the nature of the task [76]. For example, in motor imagery features can be extracted through various frequency bands, whereas in cognitive-based ERPs, properties in the time domain may prove to be more useful.

There are numerous methods reported in literature regarding the general category of feature extraction/selection. These include features derived by mutual and other information-based approaches [77, 78, 79], Fourier and other transforms [80, 22], power spectral density [81] and Fisher and other ratios [82, 83]. Selecting between features has also been approached in the literature [74, 84, 85, 86].

Electrode selection could also be included as a part of feature extraction. It is common for a single electrode not to be sufficient for the production of accurate results. Also, it may be the case that not all electrodes capture the events taking place. Moreover, the electrode location can be used as a feature on

its own, for instance to provide information regarding the laterality of the imagined movement. Selecting a number of electrodes to use could be done empirically and systematically. Common spatial patterns is an algorithm widely used for this purpose [87].

A feature vector is usually used to denote the resulting features that are passed to the classifier. In literature it has been noticed that multiple features can boost the performance of the classification [88, 89]. Features as such are referred to as combined features and they can be a simple concatenation of different features, weighted sums or even complicated manipulations of sets of features.

2.4.1 Independent and Principal Component Analysis

Two of the most commonly used tools of ERP analysis are the Independent and Principal Component Analysis (ICA, PCA respectively), methods widely used due to their power of decomposition of ERPs to components by the use of different sources [90, 91]. Essentially both methods are transforms, commonly used to produce features that are better separated than the original signal. Hence, in terms of feature extraction, the methods are applied to each epoch and a set of transforms is extracted. From the set of ICA/PCA features, a simple selection policy could be used in order to single out the features that can differentiate different responses better. Thus, the decomposition of the epochs into components that are separable from each other helps the classification process, especially when the machine is trying to linearly separate the responses (using an SVM with a linear kernel, for example).

The method has shown promise in tackling the issue of blind source separation [92, 93] for the purpose of identifying P300 components, provided it is performed on averaged ERPs, as a low signal-to-noise ratio will jade the results of the methods. Although the element of blind source separation remains a problem, it has attracted the attention of the research community, with further research being conducted [94].

One of the possible pitfalls while attempting to yield physiologically meaningful results with these methods is the likelihood of deriving significant relationships based on ERP components [95, 96]. Such analysis has been used for the habituation of the P300 wave and further hypotheses [97, 98]. Although there is no guarantee that the ICA/PCA components will have a physiological meaning, it is not unlikely that some may.

2.4.2 Common Spatial Patterns

Another major transform frequently used in BCI research are the common spatial patterns (CSP) [87]. The method produces sub-components that are separated maximally in terms of variance. In essence, this technique uses filtering to establish the best sub-component to use for the purposes of classification. In multi-class BCIs it is very common for the classes to be intertwined in the feature space and with the help of CSP these features can gain the separability required for the classification process to be accurate.

2.4.3 Eigenbrains

A recent representation of EEG signals proposed by [95] uses a different transformation. The method constructs a model of the brain, such that neurons are represented by masses and connections as springs.

The interaction between electrodes is represented as weights in a matrix w_{ij} where i, j are the sources (electrodes) and s_{ij} the voltage difference between sources, hence for two distinct electrode locations their interaction is described as $w_{ij} = 1/(0.001 + f(s_{ij}))$. The $f(s_{ij})$ function as reported by [95] can be the mean absolute error or the root mean squared value of s_{ij}. The eigenvectors of the matrix w_{ij} have been named eigenbrains.

Compared to other methods such as PCA and ICA, eigenbrains are shown to be competent and provide a clearer depiction of relationships between electrodes.

2.4.4 Filtering

Filters in signal processing cover the time, space and frequency domains. Their characterisation refers to the functionality of the filter and the nature of the data. Two major categories of filters can be used to separate the majority of filters with respect to data: discrete/digital and continuous/analogue. Filters are further distinguished in terms of linearity, time-variance, usage of external power (e.g. amplification) and impulse response [99]. Filtering is extensively used in signal and image processing to separate noise from valuable data or even to remove unwanted correlations within the data.

EEG recordings are a product of oscillatory and excitatory superposition, both from inter-cranial processes and external influences. One of the most common influences one may find in the recording is that of the power mains, which introduce a steady oscillation. Filtering has been extensively used in BCI research and literature on filtering techniques is continuously being refined for specific BCI types (e.g. [100, 101, 102]).

Band-pass filtering is one of the most commonly used filtering techniques in feature extraction since it provides a way to target specific bands. Filtering can be used to estimate the derivative of a time series and has been introduced in literature as an edge detection approach [103]. This technique has been applied to the analysis of medical images such as for the extraction of retinal vessels [104, 105].

2.4.5 Wavelet Analysis

Wavelets have been used in the analysis of the EEG by various researchers; either as a feature or a pattern recognition method, wavelets show promising results in various applications [106]. A wavelet is best described as a short-term oscillation which begins with zero amplitude and ends similarly. An example of a wavelet is the widely known *mexican hat*. The convolution of the wavelet with a signal can show areas with high correlation between the two. In literature the term “mother wavelet” is commonly used to denote the prototype wavelet. By changing the shift and/or scale parameters of that mother wavelet, one can generate an infinite amount of children wavelets. Although discrete wavelets exist, the majority of EEG analysis is conducted by continuous wavelets transforms (CWT) [107].

The mother wavelet, depending on its definition, could provide results similar to the Fourier transform or differentiation. For example, the mexican hat has been used in the identification of the extrema of a signal, i.e. its maxima and minima. The definition of the "mother wavelet" can be changed to indicate the behaviour of the 1st derivative. Since the CWT is in essence a convolution, similar results can be obtained by using filtering. For instance, when the synthetic data are filtered using the first order derivative of the Gaussian the resulting signal is near identical to the differentiation process.

Wavelets in BCI research have been used in different ways. Many researchers use the CWT in order to generate features which can later be used to classify ERPs [20, 21, 22, 23, 24]. The versatile, adaptable and smoothing nature of wavelets is valuable in EEG analysis. In the literature various researchers have published single trial, adaptive [18] and asynchronous [19] BCI systems based on wavelets.

2.5 Averaging

The possible applications of BCIs are constantly increasing but in their current state they are limited in terms of speed and accuracy [44]. A reasonable cause of the error could be that ERP variants were elicited during the test or training phase of the experiment, which may heavily affect the classification process. These variants, which were shown in [96] to differ from the output of the average, can be important; however they cannot be captured by the traditional averaging process [52].

This masking of ERP variants can partially be attributed to the fact that the signal to noise ratio (SNR) of the ERPs is very low. Hence, a large sample is needed by the average in order to produce an ERP of recognisable shape. The presence of noise and the limitations of averaging (and any variants) are known and have been investigated in [108, 109, 110].

Shape and latency characteristics are very difficult to retrieve for individual trials. Most ERPs are hidden within the brain's ongoing activity, which can severely deform their characteristics. Variations of the averaging method such as weighted averaging and probabilistic averaging [111, 112] have been introduced to minimise the deformation of the averaged ERP due to the presence of out-liers. Still, there is no averaging method currently available that could extract or describe any variability among the ERPs in an experiment.

Although optimistic, it would be beneficial to capture any distinct variability of the shapes, latencies and amplitudes of ERPs, that is, if such variations manifest during the experiment. A powerful motivation in this research are the results of [96]. The authors have shown that variants of the same ERPs could be elicited within an experiment. Moreover, the authors contemplated the use of traditional averaged EEG responses due to the keyhole effect of the averaging process, i.e. identifying the common variation between signals but destroying any non-common variation. This effect helps uncover common characteristics between ERPs but it does so at the expense of minor, possibly valuable, variability.

2.6 Clustering

Clustering of EEG signals has been investigated before; in [113], clustering was used to allocate similar responses to groups using recordings from monkeys. The majority of non-invasive BCIs has been employed in topographical analyses of EEG sources [114, 115]. Their programme of research is concerned with the classification of different ERP components (e.g. the N100, N200, P100 and P300 components) and their relations with neural network functionality. In our line of research we aim to use clustering to identify similar mental processes between different trials that elicited the same ERP.

2.6.1 Distances and Similarity Measures Between Feature Vectors

There are many clustering criteria; the ones commonly used in the literature are based on non-parametric similarity tests, distances from centroids, and information theory measures [116, 117]. Since the number of samples taken by the EEG are countable, the discrete versions of distances will be used.

n-Euclidean is the multidimensional version of the geometric distance metric evaluated as $d_{ED}(p,q) = \sqrt{\sum_{i=1}^{n}(p_i - q_i)^2}$, for i being the elements of a feature vector.

Manhattan is a block distance between points; it is calculated by $d_{Man}(p,q) = \sum_{i=1}^{n} |p_i - q_i|$

Kolmogorov-Smirnov is a non-parametric test between two empirical cumulative distribution functions with test statistic $D_{n,n'}(p,q) = \sup |q_{ecdf} - p_{ecdf}|$, where $X_{ecdf}(x \in X) = \frac{1}{n}\sum_{i=1}^{n} \begin{cases} 1, & if x_i \leq x \\ 0, & otherwise \end{cases}$

Wilcoxon rank sum is a non-parametric test between *paired* data with test statistic $W(p,q) = |\sum sgn(p_i - q_i) r_i|$, where $sgn(p_i - q_i)$ is the sign of the result and r_i is the rank of the difference when sorted, with $r = 1$ being the smallest.

2.6.2 Comparison Methods for Groups of Feature Vectors

Kruskal-Wallis is a non parametric test for homogeneity between groups. If the group is non homogeneous, then the method fails to identify which elements of the group are causing the homogeneity to fail.

Centroids of groups and distances between them could be used to estimate how far apart groups are from each other. By selecting any of the proposed distance or similarity measures a new Centroid-based method could be generated.

Temporal Evaluations are based on a similarity measure. They apply the similarity method temporally, i.e. at each given time, hence evaluating the temporal similarity between members of two groups.

2.7 Considerations on Handling Event Related Potentials

Specifically, in P300-based BCI literature it is common practice to perform a series of trials to compensate for the high noise levels where the ERP is hidden, e.g. [118, 7]. There are current studies which

evaluate performance of classifiers using single trial methods and low trial methods (typically < 3) but their accuracy is easily affected by the rate of stimulus presentation [44]. The methods used to improve the signals, as discussed in the previous sub-section, mostly depend on variations of averaging, e.g. [119, 111].

Averaging distorts the original signal, revealing its main components but destroying any other minor variability. As shown in [120] and discussed in [121], stimulus-induced rhythmic EEG activity, which is not phase-locked to an event, is also contained within individual trials; this activity is not detected by averaging. Time domain processing techniques such as linear averaging remove non phase-locked EEG activity by phase cancellation. Existent variants of the mean are prone to outliers and heavy tails, e.g. [109].

In the work done by [122], spectral analysis has been criticised on its use in the analysis of evoked potentials. The authors comment on the validity of spectral analysis and suggest that although brain dynamics are commonly analysed by linear systems (such as the fast Fourier transform), the brain is a highly complex, non-linear system. Moreover, the authors point out the need for multiple trials for the averaging process and also showed that the spectral power curves for the input signal were vastly different when compared to the output from the above methods. The authors conclude suggesting that the evoked potentials hidden within other brain activity cannot be estimated by the results of such analysis.

Regarding the interpretation of averaged ERPs, suggesting that they approximate the true ERPs is wrong [96]. The average of a collection of epochs, thought to contain an ERP, is no more than an estimator of any time-locked events that occurred throughout the epochs. The method's output cannot be regarded as an ERP itself. The authors suggested that the current averaging methods fail to provide inference on the true ERP. Thus, inferring to a P300 response using an averaged response could be erroneous. It is difficult to assume that there is a distinct ERP or that target stimuli elicit the same ERP without a pre-averaging classification of ERPs. Yet in literature averaged ERPs are widely used as a method of finding ERPs and evaluating their physiological meaning.

Recent research has shown that different averaging methods used for the same data-set yield different results; a binning method was proposed to increase the resolving power of ERP averages, and the results reported showed significant variations in the components of resulting ERP epochs [96].

2.8 Paradigm and Data Acquisition

The data-set used in the thesis was acquired using a P300-based BCI mouse [123, 96, 124, 125, 126], with a periodic stimulation protocol. In the paradigm, 8 circles were distributed around the centre of the screen in front of the participants (see Figure 2.1, page 27). The circles flashed in a sequential manner in different colours: green and red. The participant was instructed to focus on the circle corresponding to the desired direction of pointer motion and to mentally name the colour of its flashes to strengthen ERP responses. Non-flashing circles were coloured grey and the duration of the flash (stimulus) was 100 ms, with no delay between flashes. The experiment was carried out with 16 participants. All participants were comfortably seated and asked to minimise movements and blinks during the experiment. Data

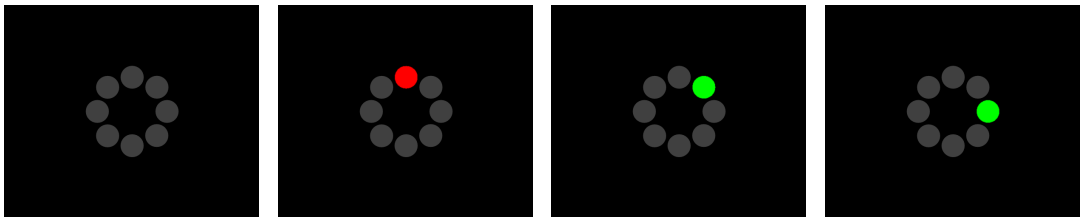

Figure 2.1: Stimulus presentation of the P300 based BCI mouse. From left to right: initial display and three sequential stimuli.

were acquired from the Active-II BioSemi system at 2048 Hz, using 64 electrodes, allocated according to the international 10-20 system, while measurements from the earlobes were used for referencing.

Chapter 3

Derivatives and Features

The majority of BCI literature focuses on amplitude-based features since the P300 is characterised as an intense wave captured throughout the surface of the scalp. In BCIs, the existence of an ERP is not the only information that is sought after; differences between the characteristics of ERPs, i.e. their source, latency, and propagation, may provide useful information. The amount of information a BCI can extract relevant to a task is a major factor of its applicability.

3.1 Theoretical Investigation of the Derivative

Theoretically one can show that the variance of the derivative is twice that of the amplitudes. Let us assume that the equation of the amplitude of the EEG can be given by:

$$v_i(t) = s_i(t) + n_i(t)$$

where i is class, e.g. P300 ERPs, s is signal and n is noise, i.e. any influences from the working of the brain coming from deeper regions, are assumed to be **additive** and **uncorrelated** to s_i. Then the standard deviation of the amplitude can be calculated as:

$$\sigma^2[v_i(t)] = E[v_i(t)^2] - (E[v_i(t)])^2 =$$

$$E[(s_i(t) + n_i(t))^2] - (E[s_i(t) + n_i(t)])^2 =$$

$$E[s_i(t)^2] - (E[s_i(t)])^2 + E[n_i(t)^2] - (E[n_i(t)])^2 + 2E[s_i(t)n_i(t)] - 2E[s_i(t)]E[n_i(t)] =$$

$$\sigma^2[s_i(t)] + \sigma^2[n_i(t)] + 2E[s_i(t)n_i(t)] - 2E[s_i(t)]E[n_i(t)] =$$

$$\sigma^2[s_i(t)] + \sigma^2[n_i(t)].$$

Similarly for the derivative, using the forward approximation (forward difference) and the same assumptions, the standard deviation is:

$$\sigma^2[d_i(t)] = E[d_i(t)^2] - (E[d_i(t)])^2 =$$

$$E[(s_i(t+1) - s_i(t) + n_i(t+1) - n_i(t))^2] - (E[s_i(t+1) - s_i(t) + n_i(t+1) - n_i(t)])^2 =$$

$$E[s_i(t+1)^2]+E[s_i(t)^2]+E[n_i(t)^2]+E[n_i(t)^2]$$

$$-2E[s_i(t+1)s_i(t)]+2E[s_i(t+1)n_i(t+1)]-2E[s_i(t+1)n_i(t)]$$

$$+2E[s_i(t)n_i(t)]-2E[s_i(t)n_i(t+1)]-2E[n_i(t)n_i(t+1)]$$

$$-((E[s_i(t+1)])^2+(E[s_i(t)])^2+(E[n_i(t)])^2+(E[n_i(t+1)])^2$$

$$-2E[s_i(t+1)]E[s_i(t)]+2E[s_i(t+1)]E[n_i(t+1)]-2E[s_i(t+1)]E[n_i(t)]$$

$$+2E[s_i(t)]E[n_i(t)]-2E[s_i(t)]E[n_i(t+1)]-2E[n_i(t)]E[n_i(t+1)])=$$

$$E[s_i(t+1)^2]-(E[s_i(t+1)])^2+E[s_i(t)^2]-(E[s_i(t)])^2+E[n_i(t)^2]-(E[n_i(t)])^2+E[n_i(t+1)^2]-(E[n_i(t+1)])^2$$

$$-2E[s_i(t+1)s_i(t)]+2E[s_i(t+1)n_i(t+1)]-2E[s_i(t+1)n_i(t)]+2E[s_i(t)n_i(t)]$$

$$-2E[s_i(t)n_i(t+1)]-2E[n_i(t)n_i(t+1)]+2E[s_i(t+1)]E[s_i(t)]-2E[s_i(t+1)]E[n_i(t+1)]$$

$$+2E[s_i(t+1)]E[n_i(t)]-2E[s_i(t)]E[n_i(t)]+2E[s_i(t)]E[n_i(t+1)]+2E[n_i(t)]E[n_i(t+1)].$$

By assuming that the noise is **white**, the above becomes:

$$\sigma^2[s_i(t+1)]+\sigma^2[n_i(t+1)]+\sigma^2[s_i(t+1)]+\sigma^2[n_i(t+1)]-2E[s_i(t+1)s_i(t)]+2E[s_i(t+1)]E[s_i(t)].$$

If one sample corresponds to a **very short amount of time**, then

$$\sigma_i^2(t+1)\cong\sigma_i^2(t).$$

Thus,

$$\sigma^2[d_i(t)]\cong 2\sigma^2[s_i(t)]+2\sigma^2[n_i(t)]-2E[s_i(t+1)s_i(t)]+2E[s_i(t+1)]E[s_i(t)].$$

Now, given that a **baseline removal process** was applied it is reasonable to assume that s_i **has zero mean**, hence

$$\sigma^2[d_i(t)]\cong 2\sigma^2[s_i(t)]+2\sigma^2[n_i(t)]-2E[s_i(t+1)s_i(t)].$$

The derivative in terms of signal processing acts as a high-pass filter. From the equations above, one can understand that if the s_i is a low-band signal then the expectation of the product of the signal at adjacent samples is positive, thus reducing the variance of the derivative. On the other hand, if the signal is high-band then the expectation of the product of the signal at adjacent times is negative, which consecutively adds to the variance. If the expectation of the product is zero, then it is clear that the information from the derivative is similar to that of the amplitudes, while the variance of the derivative is twice that of the amplitudes.

In theory, the EEG can be re-constructed from its derivative, thus suggesting that the derivative holds important information about the amplitude. The increase in variance for high-band signals shows that there is more information hidden within the derivative rather than the amplitudes. One cannot be certain whether this increase in information contributes to the ERP or is irrelevant. Since the focus of researchers has been the identification of the ERP and amplitudes have produced reasonably good results, literature on exploiting derivatives is scarce. It is still unknown whether the increased variance of the derivative can provide more valuable information than the amplitudes.

3.1.1 Empirical Validation of the Theory

To test the theoretical analysis, firstly synthetic signals are created as shown in Figure A.1, page 71. Using a sinusoidal model, the signal (a) is added to white noise, thus creating a first set of synthetic signals to test (b). Since the construction of the synthetic data (a) is very close to the definition used in the theoretical analysis, it is expected that the standard deviation is going to behave as predicted by the theory. In this case, the amplitude is expected to have half the variance of the derivative since for the synthetic data, $s_i(t+1) \cong s_i(t)$.

In the synthetic data used to produce Figure A.2, on page 72, there is no filtering performed. On page 72, Figure A.3 shows the effect on standard deviation having low-passed the data. The standard deviation of the resulting synthetic ERPs is now clearly not abiding by the theory. The same behaviour can be seen in real data (described in section 2.8, page 26).

In the real data we can expect three responses;(1) a P300, which is elicited by the stimulus, (2) a SSVEP which is the direct outcome of the periodic stimulus presentation and finally, due to colour changes and the expectation of the participant, there is a chance that (3) an MMN is also elicited. Figure A.4 (page 73) shows the standard deviation of both the first derivative and the potentials, using the raw data.

The standard deviation of both the voltages and the derivative is very similar to that of the low-passed synthetic data results. After closer investigation on the hardware used in the experiment, the digitisation process has been found to make use of a low-pass filter. By introducing a notch filter and clearing the data from the SSVEP component the standard deviation of the potentials drops significantly (Figure A.5, page 74). The removal of further contamination such as MMN, respiratory and cardiac cycles requires information which was unavailable at the time of this research.

3.2 The First Derivative as a Feature

EEG recordings can provide data with high temporal resolution. The core of a BCI's operational capacity is the classification part, commonly tackled by machine learning. In the process of teaching the machine to recognise or distinguish between classes, there is always the challenge of over- and under-fitting the data-set. The majority of BCI experiments split their recordings into training and testing sets. While training the classifier, one must have at least as many examples as the amount of information passed to the classifier (features).

Since the sampling rate of the recording in time is usually more than the number of ERPs one can capture during a training experiment, the machine learning element of the system cannot learn properly, so the research community has developed several methods to mitigate this. Sub-sampling is one of the most commonly used methods for simplicity. Window selection is an approach which aims at both the dimensionality reduction and the quality of the feature vectors. Another way researchers go about this issue is splitting the feature set into different classifiers which are later used in deciding the classification outcome.

The quality of a feature can be estimated in terms of the resulting accuracy and success of the clas-

sification process. Ultimately, a perfect set of features in the case of a P300 BCI would consistently identify the existence of the P300 when stimulus is presented. Hence, for the purpose of this comparison an ensemble of linear SVM classifiers is used. To ensure that each classifier performs adequately, a grid search was conducted using the mutual information criterion, in order to select the best possible C parameter.

3.2.1 Window Selection for P300 Features

Figure A.6 shows the theoretical AUC calculated by using the assumption that the temporal distributions of dV are normally distributed, that is, the distribution of the dV at each sample. By assuming such a premise one can obtain per-sample theoretical AUCs (area under the receiver operating characteristics curve) by calculating the overlap of the probability density functions for dV(T) and dV(NT) epochs (T if the stimulus was the target of interest and as NT otherwise).

In practice, most researchers follow the same feature extraction methods for the detection of a P300: epochs of 800ms and sampling rates of 32-64Hz. As seen in Figure A.6 (page 75), the amplitudes of the P300 differ between Ts and NTs in a wide centre part of the epoch. Hence, sub-sampling from the initial recording frequency (2048Hz) to lower ranges not only improves the difference by removing samples that contain cumbersome oscillations, but also reduces the number of features that are correlated. Contrary to the above, theoretical AUCs based on the first derivative indicate dV have differences in a narrow window at 300-400ms after stimulus presentation (see Figure A.7, on page 76). The time period indicated by this process is compatible with the ERP at hand (here believed to be a P300, i.e., an increase in activity 300ms after stimulus onset).

In this case, excessive sub-sampling and a wide window would hinder the classification performance. Furthermore, the usage of the whole epoch as a feature will hinder classification performance since the amount of valuable information can only be derived from a small set of samples. The above can be confirmed by looking at the AUCs calculated using a moving window and a real classifier. As seen in Tables 3.1 and 3.2, as the window moves from 0ms (after stimulus onset) to 800ms in steps of 100ms, the AUCs for the derivative are attaining their maximum at only one step. As expected, in the same test, the ERP itself shows increased AUCs in most of the 100ms windows, including that of the derivative.

It is of interest that the classification here, for different participants, is better at varying time intervals. Although the probability of partially capturing the P300 in both windows (200-300ms and 300-400ms) exists, it is also possible that participants with high AUCs during the 200-300ms time window actually elicited a MMN rather than a P300.

3.2.2 Sparse Window Selection using Amplitude Features

The generation of the P300 wave consists of synchronised firing of neurons. When the firing of a neuron is completed, a resting (or decay) period follows [52]. In between these two processes the ERP attains its maximum potency and it is usually in this region that researchers can identify the existence of an ERP.

Selecting features according to ADM is justifiable since it is assumed that the noise present in the

P	.0-.1	.1-.2	.2-.3	.3-.4	.4-.5	.5-.6	.6-.7	.7-.8
1	62.0	55.5	67.9	**72.0**	60.9	58.0	62.2	54.5
2	55.8	63.4	65.1	**80.1**	63.1	59.2	59.8	57.4
3	57.1	65.3	**75.5**	69.6	63.5	62.7	53.6	58.8
4	56.0	60.9	65.4	**69.3**	67.3	56.4	56.1	51.3
5	55.5	70.0	61.2	**73.5**	64.2	60.2	54.3	57.0
6	55.4	55.8	56.6	56.1	**57.2**	53.1	56.6	56.5
7	58.6	64.5	72.0	**73.1**	64.1	63.3	55.9	53.3
8	58.5	66.8	**69.2**	65.6	57.0	52.8	62.0	59.5
9	56.3	62.6	**70.5**	68.5	65.8	66.0	59.9	55.2
10	63.6	72.2	**75.4**	74.0	63.6	52.3	59.8	61.0
11	62.3	65.3	61.4	62.4	**71.5**	65.1	58.0	52.3
12	59.0	66.7	71.6	**78.7**	75.7	70.2	57.9	56.8
13	52.8	69.3	**72.9**	65.1	62.0	53.9	54.2	54.3
14	59.4	68.3	**82.4**	79.9	63.1	68.7	65.9	53.4
15	54.1	54.9	**66.2**	64.8	62.9	63.2	52.7	50.2
16	62.7	57.2	56.2	61.4	58.3	60.2	**64.5**	56.5

Table 3.1: Area under the receiver operating characteristics curve (AUC) as a percentage resulting from an 8-fold cross-validation using an ensemble of 3 SVMs and a moving window of 100ms. Results for dV.

P	.0-.1	.1-.2	.2-.3	.3-.4	.4-.5	.5-.6	.6-.7	.7-.8
1	69.8	74.1	82.2	**89.9**	85.1	85.1	75.8	70.0
2	68.6	73.0	77.7	**85.5**	82.0	77.3	68.7	59.6
3	67.6	74.0	**85.3**	84.3	71.7	67.0	62.2	63.3
4	69.3	74.8	77.1	81.7	**82.4**	72.6	67.1	64.3
5	77.5	79.6	76.1	**84.2**	76.9	74.4	71.8	69.4
6	65.6	68.2	70.3	**77.8**	75.4	68.9	64.1	64.9
7	70.9	74.7	79.4	**81.9**	79.2	77.4	68.2	68.1
8	71.7	79.4	84.7	**87.7**	84.6	82.2	81.6	74.3
9	69.5	74.6	81.9	**84.1**	82.9	79.7	69.1	65.8
10	76.6	86.7	**88.2**	87.8	81.3	77.8	72.2	70.7
11	65.8	69.7	74.6	77.7	**80.2**	78.6	70.3	64.7
12	72.0	77.6	84.6	**87.9**	85.7	79.8	65.5	63.9
13	72.5	81.0	86.2	**87.1**	77.7	71.5	67.3	66.2
14	75.9	80.0	91.5	**92.0**	85.3	82.0	75.2	65.6
15	60.8	64.6	**77.9**	76.2	74.8	75.2	64.1	60.4
16	70.8	67.8	69.2	78.2	74.4	78.2	**79.3**	74.0

Table 3.2: Area under the receiver operating characteristics curve (AUC) as a percentage resulting from an 8-fold cross-validation using an ensemble of 3 SVMs and a moving window of 100ms. Results for V.

data in each class is approximately identically distributed. Thus the ADMs correlate well with the classification accuracy provided by the selected feature (used in isolation). The application of this feature selection method to each channel independently ensures that feature vectors contain an equal number of samples from every recording site. This was expected to give more robustness to the resulting classifiers as noise at distant sites is less likely correlated, and more likely to cancel out. Furthermore, we made available the whole epoch to the selection method as ERPs manifest themselves at different times at different sites. Also, dV may pick up useful information from both the polarisation and depolarisation fronts of an ERP, which are, perforce, before and after the traditional locations of known ERPs.

Figures A.8 and A.9 (pages 77, 78) illustrate the behaviour of our sample selection algorithm for each channel, participant and EEG representation (V and dV). For each channel and participant we selected the 24 samples with the highest ADM (for either V or dV, estimated via cross-validation) and plotted them as small segments in the figure. The vertical position of each segment represents the time within the epoch at which the corresponding sample was acquired (see temporal scale next to the label "P1" in the top left corner of each panel).

If we look at each participant individually, the figure shows that there are significant variations in choice of optimal samples across electrodes. For V, with some participants, one fixed time window corresponding approximately to the P300 ERP would have worked well across all channels (e.g. P12 or P15), but for many a fixed time window would have been sub-optimal (e.g. P5, P11, or P13). For the latter, our sample selection method can provide significant advantages. Also, if we look at dV, we find that no single window would have been ideal, as at the very least both fronts of the P300 would need to be captured, and these are often quite far apart. Also, we see much more variability across channels as to the temporal location of the optimal samples. If we then look across participants, we see that really, any fixed set of time windows would have provided sub-optimal results for both V and, even more so, dV.

We should note that in the figure we have distinguished between the best 12 samples (in red) and the following 12 best samples (in blue). The reason is that we wanted to see to what degree the information contained in the second half of the selected features is different from that provided by the top half and adds to the classification accuracy. As we can see in the figure, the red and blue bars tend to cluster more in the V panel than in the dV panel. Indeed, for dV we see that with only 12 samples per channel, for many participants and channels one can only capture either the positive or the negative fronts of the P300, but not both.

3.2.3 Filtering and Sub-sampling

As seen in the introduction of this section, the use of filtering can be beneficial in terms of reducing the variance of the recordings. The notch filter used had the property of extracting a specific oscillation at a certain phase. Hence, even if information were extracted from the recordings, that information would be specific and could be accounted for. Although the notch filter is very helpful in such situations, the recordings hold information that is derived from a wide range of frequencies. Depending on the BCI's nature, one might want to separate, extract or discard these frequencies. The band-pass filter is one of the most commonly used for the extraction of frequency information from the recordings and clearing

them of unwanted disturbances such as electrical influence from the mains.

The amount of sub-sampling that can be performed without loss of information is bound by the sampling rate of the recording, the filtering that has been applied prior to the sub-sampling process and the length of the labels. The lower cut-off point can be calculated as twice that of the lower cut-off frequency of the filter. For example, if the data had been low-passed at $15Hz$ one could sub-sample as low as $30Hz$. The length of the labels could also limit the amount of sub-sampling that can be safely performed because they might overlap. In our data, sub-sampling to rates lower than $256Hz$ will create overlaps in the labels of the responses, hence for comparing V to dV, this sampling rate or a higher one should be used. Regardless of the label overlap, when using amplitude features it is common to sub-sample the original recordings down to $32Hz$, $64Hz$ or 128; such sub-sampling diminishes the need for a feature selection process since the whole epoch could be used. Just as the amplitude has a range of sampling rates that perform better in classification, so the derivative should have a range of sampling rates that better fit the information it holds.

Reducing the amount of data at hand is usually tackled initially by sub-sampling before using more sophisticated feature selection methods. Aliasing is one of the most common pitfalls of sub-sampling. Although there is a wealth of data from the recordings, an ERP such as the P300 will still only be observable for a short period . To put things in perspective, an $800ms$ epoch will have an ERP which lasts about $300-400ms$. According to the labels in our data the epoch accounts for one T stimulus presentation and seven consecutive NT stimuli. Hence, if the recording was first sampled at $2048Hz$, the resulting epoch would contain 1638 samples (approx. 205 samples for each stimulus, T or NT). This translates to approximately $615-820$ samples where the ERP is discernible. A very low sampling rate such as $32Hz$ would only have $9-12$ where the ERP would be captured.

As regards classification, if an experiment produced 24 ERPs, the feature vector should have approximately 24 features that could be selected from the total number of samples per epoch in order to account for over- and under-fitting of the classifier. Since the feature vector length depends on the sampling rate, a simple feature selection process is carried out. For ranking the samples, the ADM was used as previously described. Hence 24 samples were selected per electrode (all electrodes used) as features to train a simple LDA classifier. A more sophisticated classifier, such as an SVM ensemble trained to use mutual information coefficients and a grid search for the optimal C value, would produce higher classification performances but at the cost of increased training times. Since here we are essentially conducting a grid search of the size $(\texttt{sampling rates}) \times (\texttt{filtering upper cut}-\texttt{off}) \times (\texttt{amount of derivative features}) \times (\texttt{cross}-\texttt{validation folds})$ per participant, we have to train the machine $7*3*4*8*16 = 10,752$ times for all participants. Using an ensemble with a grid search to optimise its performance would increase that number dramatically, so if one were to search for each SVM in an ensemble of three, the optimal C value out of ten possible results in $10752*3*10 = 322,560$ training rounds.

Mixing features from both the amplitude and the derivative could increase the performance of classification. In Figures A.10—A.12 (pages 79—81), the results of a concatenated feature vector are displayed for different sampling rates and cut-off frequencies. The varying components of this exploration is the

upper limit of the filter, set to vary between $5Hz$, $15Hz$ and $30Hz$, and the sampling rate. Initially, the experiment considered a feature vector derived entirely from the amplitude, while every repetition of the experiment switched V-features with dV-features, and so the final experiment concluded with a feature vector completely comprised by dV-features.

#dV	5Hz	15Hz	30Hz
5	-	-	-
20	0.0017	0.0022	0.0042
24	0.0033	0.0017	0.0041

Table 3.3: Significance testing using the 1-tail Wilcoxon test of AUCs from V-based feature vectors and dV concatenated features (when the #dV is 24 the feature vector is completely comprised of dV features) at different filtering strengths for $2048Hz$ sampling rate

Reducing the resolution of the epochs follows an increase in the classification accuracy of both the derivative and amplitude feature vectors. The general tendency of the participants is in favour of the $15Hz$ filter upper limit. Also, it is clearly displayed that using a high cut-off frequency ($30Hz$) hinders the classification of the derivative-based features for most participants, irrespective of the sub-sampling performed. Furthermore, low sampling rates show that both V and dV can yield high classification performances; as the sampling rate increases, it is evident that dV features are more robust than V. Finally, the results suggest that selecting features from high sample rates significantly reduces (Table 3.3) the classification performance for V when compared to dV.

3.3 Combining Amplitude and Derivative Based Selection of Features

The slope as a feature is complementary with the amplitudes, i.e. the samples that seem to differ greatly from the slope are not the same as the ones reported by the amplitudes (see Figures A.8, A.9) (pages 77, 78). In the figure we show the absolute difference in the medians, according to target direction flashes and non-target flashes. For some participants, the complementary nature of the amplitudes and the derivatives as features can be clearly seen (e.g., P12).

The following experiment aims to show the impact to amplitude classification, thanks to the inclusion of samples selected by the slope. Hence, instead of selecting a sparse window solely based on the ADM of the amplitudes, here the window for the amplitudes is selected using both ADMs, derivative-based and amplitude-based.

Although previous sections have stated that sampling lower than $256Hz$ the labels from the experiment will start to overlap, the sampling rates chosen were $64Hz$ and $128Hz$. Here we aim to push the boundaries of classification performance and as suggested beforehand, using amplitude features one may further down-sample at an epoch level. Since differences between Ts and NTs in terms of amplitude are widely spread over the duration of the epoch, further sub-sampling would only remove samples holding

information which is also held by neighbouring samples. Conversely, the derivative has differences in a much shorter window than that of amplitudes. Initially, 24 samples from each electrode location were picked (out of 51 and 102 samples for $64Hz$ and $128Hz$ sampling rates, respectively) using an amplitude-based ADM. The number of amplitude selected samples used in the final feature vector were thereafter gradually decreased by replacing them with derivative-based ADM selected samples, until all samples were selected using the derivative. The exchange process employs a worst out, best in policy according to the value of the ADM.

An ensemble of 3 linear SVMs was used for classification. According to the labels (T or NT), the training set was split evenly amongst these classifiers. A grid search was conducted and the mutual information coefficient was used to estimate the best possible C parameter for each SVM in the ensemble. The classification process consisted of an 8-fold cross-validation. Epochs were mixed in such a way that each training set always contained epochs from all directions. The mean of the ensembles output was used to compute the Receiver Operating Characteristics (ROC).

During training of the ensemble, participant 6 was problematic when the upper bound of the filter was at $5Hz$ and hence removed from this analysis. Table 3.4 shows the configurations that provided the best AUC for each subject, whereas 3.5 shows the worst possible. The classification accuracy, described by the AUC, is increased for each subject when more samples are selected by the derivative. Significance testing on the AUCs show that by using higher sampling rates, the derivative-based ADM selection yields better classification results when using amplitude features (see table 3.6). Depending on the subject, the slope shows that it can either completely replace the use of amplitude differences in the selection process or enhance it.

Although the number of derivative samples is not the same for all participants, including derivative selected samples in feature vectors seems to have a positive outcome. Also, the majority of the participants reached their highest classification rate when the feature vectors contained combined information from both the slope and the amplitudes. Not surprisingly, the configurations yielding the lowest possible AUCs resulted from data being filtered up to $5Hz$. This confirms that a poorly preprocessed ERP contains variation that is not beneficial to classification, irrespective of the sample selection method.

The above investigation of sample selection based on the derivative of the ERPs has been shown to benefit a portion of the participants. Table 3.6 summarises the results in terms of statistical significance across participants. The increase in AUC is significant at a 5% confidence limit and from the results stated above few things can be deduced. Firstly, the use of derivatives can be beneficial in terms of increase in AUC when used to select samples for amplitude-based feature vectors. Secondly, not all participants are greatly benefited by this selection process; in such cases the samples selected by the slope certainly did not have a major impact on the AUCs. It has also been shown that filtering and sampling rate combinations can have a massive impact on classification performance. In our understanding, derivative selected samples are more robust than the amplitude selections which focus on the resting period of the ERP. Finally, the combined features proved to be rewarding with the majority having an inclination towards slope selected samples.

Participant	SR(Hz)	UC(Hz)	dV	AUC
1	64	15	20	91
2	64	15	20	90
3	64	15	12	86
4	64	15	24	87
5	64	15	20	86
6	64	30	20	88
7	64	15	0	90
8	64	15	0	88
9	128	15	24	90
10	64	15	12	83
11	64	30	24	90
12	64	15	12	90
13	128	15	20	95
14	128	30	12	84
15	64	15	20	84

Table 3.4: Table containing the best possible configurations for sampling rate (SR), filter upper cut-off frequency (UC) and number of derivative samples selected (dV).

Participant	SR(Hz)	UC(Hz)	dV	AUC
1	64	5	0	70
2	128	5	24	72
3	128	5	0	72
4	128	5	0	75
5	128	5	0	73
6	128	5	24	78
7	128	5	24	79
8	128	5	12	77
9	128	5	5	80
10	64	5	24	78
11	64	5	20	77
12	128	5	0	76
13	128	5	0	78
14	128	5	20	71
15	128	5	0	75

Table 3.5: Table containing the worst possible configurations for sampling rate (SR), filter upper cut-off frequency (UC) and number of derivative samples selected (dV).

dV	5Hz	15HZ	30Hz
5	0.2562	0.2062	0.3719
12	0.2562	0.1164	0.2181
20	0.3264	0.0936	0.1427
24	0.3719	0.1084	0.2834

dV	5Hz	15HZ	30Hz
5	0.1248	0.1164	0.2562
12	0.1623	0.0685	0.1008
20	0.0803	**0.0306**	0.0742
24	0.1523	**0.0488**	0.1834

Table 3.6: Table containing p-values from the Wilcoxon one-tailed test between AUCs resulting from amplitude only against AUCs of combined features; dV denotes the number of samples selected according to the derivative. Results shown across participants for the different filter upper cut-off frequencies, number of features selected by the derivative and sampling rates (left 64Hz, right 128Hz).

3.4 Comparing the EEG to its Time Derivative

According to previous investigations on filtering, in order to increase the signal to noise ratio and remove any contamination from the mains (and any other high frequency artefacts), the data were band-passed between 0.15 Hz and 15 Hz. As suggested, the periodicity of the flashing stimuli introduced a SSVEP. We used a Notch filter to remove the corresponding 10 Hz frequency. Despite our instructions, there were involuntary eye movements and eye blinks that introduced artefacts in the EEG. To reduce their effects, a correction was performed by using correlations and a standard subtraction method was applied using the average of the paired differences of channels Fp1/F1 and Fp2/F2 [127].

Prior to feature extraction and selection, the EEG was epoched. Windows of 800 ms starting from each stimulus onset were extracted and labelled as T if the stimulus was the target of interest and as NT otherwise. Epochs were then down sampled at 256 Hz (resulting in 205 samples per epoch). A baseline removal process for each epoch was also used: the baseline was calculated by averaging 100 ms of signal prior to the epoch. A total of 16 recordings per participant were available, each including approximately 22 Ts and 154 NTs.

In this analysis we are interested in the performance yielded by three feature vectors formed by selecting samples from: (1) the EEG, V, (2) its time derivative, dV, and (3) their combination, VdV. If v_t is the recorded voltage at time t at a particular channel i, then the amplitude feature vector for a given stimulus is $V_i = v_0, \ldots, v_t, \ldots, v_n$. The first order derivative of this feature vector, dV, is calculated using the central differences method:

$$dV_i = \begin{cases} v_{t+1} - v_t, & \text{if } t = 0 \\ \frac{v_{t+1} - v_{t-1}}{2}, & \text{if } 1 \leq t \leq n-1 \\ v_t - v_{t-1} & \text{if } t = n \end{cases}$$

Finally the combined feature vector VdV was constructed by simply concatenating V and dV.

In both V and dV, out of the 205 available samples 24 were selected with the highest ADM between T and NT epochs in the train set, for each electrode location. For comparison fairness, the combined feature vector was constructed from both V and dV, but using only the 12 best samples from each, resulting in 24 samples per electrode.

F	**Band-pass only**					**Band-pass and Notch**				
S	**12**		**24**			**12**		**24**		
P	**V**	**dV**	**V**	**dV**	**VdV**	**V**	**dV**	**V**	**dV**	**VdV**
1	82	86	84	**88**	**88**	82	86	84	**88**	**88**
2	78	87	80	**90**	86	79	87	79	**90**	85
3	80	82	81	**84**	82	78	81	81	**84**	81
4	78	79	81	**83**	79	77	79	**83**	**83**	80
5	74	86	74	**88**	81	74	85	75	**87**	79
6	72	72	71	**74**	73	71	73	71	**74**	73
7	80	80	81	**83**	**83**	80	81	81	**83**	**83**
8	88	84	**89**	88	**89**	**89**	84	**89**	86	**89**
9	81	85	84	**87**	85	83	85	84	**87**	86
10	77	86	80	**87**	86	78	86	80	**89**	86
11	79	79	78	**80**	**80**	79	78	**80**	79	**80**
12	85	84	86	**89**	87	84	85	86	**88**	87
13	89	86	89	88	**90**	**90**	86	**90**	88	88
14	86	93	88	**95**	91	85	93	89	**95**	91
15	72	75	72	**76**	**76**	71	76	72	76	**77**
16	79	75	**82**	78	80	79	75	**80**	77	**80**
Med.	80	84	81	**87**	83	79	84	81	**86**	83

Table 3.7: AUC for each participant when using 12 and 24 samples (S) for V and dV and combining 12 samples of each into VdV. Results without the notch filter (F) are displayed in the left half of the table. Best results in each half of the table and for each row are shown in bold face.

An ensemble of 3 linear SVMs was used for classification. According to the labels (T or NT), the training set was split evenly amongst these classifiers. A grid search was conducted and the mutual information coefficient was used to estimate the best possible C parameter for each SVM in the ensemble. The classification process consisted of an 8-fold cross-validation. Epochs were mixed in such a way that each training set always contained epochs from all directions. The mean of the ensembles output was used to compute the ROC.

Classification performance was assessed using the AUC, as a percentage. As indicated above, an 8-fold cross-validation process was used and thus the average AUC across folds was used as a reliable description of classification performance. Table 3.7 shows individual AUC results using both 12 and 24 samples per electrode, selected using the ADM for V and dV, with and without the application of the notch filter. The table also reports the median AUC across participants.

From the table we can see that generally using samples of the first derivative of the EEG provides better performance than using the voltages themselves, whether 12 or 24 samples are used. However, 24

Table 3.8: Mann-Whitney one-tailed test of statistical significance of the AUCs obtained when using V, dV and VdV, using 12 and 24 samples (S), with and without notch filtering. We have omitted tests comparing configurations with different numbers of features.

	Band-pass only			**Band-pass and Notch**		
S	**VdV>V**	**VdV>dV**	**dV>V**	**VdV>V**	**VdV>dV**	**dV>V**
12	—	—	0.102	—	—	0.081
24	0.145	0.225	**0.050**	0.224	0.248	0.087

samples increase classification accuracy significantly over 12 irrespective of the representation used.

In relation to VdV we see that it somehow appears to be half way between V and dV in terms of performance. While this is reasonable, it goes against previous results obtained with a BCI matrix speller which suggested that combining the two representations was best [16, 17].

Finally, comparing the left and right sides of the table reveals that the notch filter has very little effect on classification accuracy. Visually, we found that the ERP averages (not reported) produced with notch filtering better resemble textbook type ERPs. However, because the ripples caused by the SSVEP associated with the periodicity of stimulation are present with approximately the same amplitude in both targets and non-targets, the filter makes relatively little difference to the separability of the two classes.

We have performed statistical tests to assess the veracity of these observations. Results are reported in Table 3.8. As one can see, while V and VdV are not significantly different statistically, in almost all comparisons between V and dV, dV was either significantly superior statistically or close to being statistically superior to V. This suggests that dV may be a better representation of ERPs, at least in the BCI system considered here.

3.5 Chapter Conclusions

Derivative based features are rarely used in BCIs, although at least theoretically, they contain almost the same information as the EEG itself since one can be reconstructed from the other.

In this chapter I evaluated the performance of the first temporal derivatives of the EEG for the classification of ERPs with data from 16 participants using a P300 based BCI mouse. A sample selection method was introduced (ADM) based on the temporal absolute difference between the medians of responses to target and non-target stimuli and a comparison was conducted between the classification performance of voltages, derivatives and a combined feature vector, i.e. by using both the amplitudes and their derivative.

The results indicate that the use of the derivative improves classification accuracy, both in combination with amplitudes and even more so when used on their own. Taken together with findings reported in previous literature, where derivatives were used in conjunction with PCA, our results suggest that temporal derivatives warrant further investigation as features for BCI systems.

In the future, more sophisticated methods of extracting the derivative could be used, e.g. least squares approximation, a Savinsky-Golay and even certain families of wavelets such as rbio3,1. In addition, the behaviour of transformations such as ICA, PCA and eigenbrains could be investigated to further boost the classification performance. Finally, for the purposes of this thesis all electrode locations were used, thus keeping a higher spatial resolution; future work could evaluate the relative benefits of performing source selection based on the first derivative of the EEG.

Chapter 4

Clustering

4.1 Enhancing ERP Analysis and BCIs by Clustering

It has recently been suggested [96] that binning ERPs based on user response times (before averaging) could allow one to distinguish and observe with a higher resolving power the ERPs associated with distinct mental responses to identical stimuli. However, this technique relies on the availability of a physical manifestation of a user's mental processes (a response, e.g. in the form of a key press), and could not be used in experiments where such a manifestation is not desirable or possible, for instance in the case of an enhancement-BCI.

In this chapter, we attempt to go beyond this limitation by using forms of unsupervised clustering of ERPs. The idea is that different mental processes are bound to produce different ERPs. It should therefore be possible, at least in theory, to group ERPs elicited by certain classes of stimuli (e.g. targets or non-targets) into smaller subsets in such a way that the members of each subset are likely to represent the same mental processes and *vice versa*. If this were achieved, there would be some noteworthy consequences. Firstly, it would be possible to do averages only across each subset of ERPs, thereby increasing their resolving power even without any external manifestations of mental state. Secondly, it would be possible to create distinct (and much cleaner) training sets for the range of behaviours triggered by an event as well as build a whole new form of classifiers, each specialised in dealing with a specific mental response. Thirdly, one would be able to first classify unseen ERPs as belonging to one of only a few possible mental processes, and then pass them only to suitable classifiers.

Figure 4.1 gives an example of the new possibilities made available by this approach in the analysis of mental processes. Here we imagined that there are two kinds of stimuli - targets and non-targets - and that for each there can be two mental processes taking place: either the participant is focused on the task and does it competently or the participant is daydreaming. The corresponding four kinds of ERPs elicited in these conditions are shown at the very top of the figure. The traditional analysis, which applies averaging based on external conditions (target vs non-target), is represented by the two triangles just below the ERPs. This can only produce the target and non-target averages below the triangles, which are blurred mixtures of the "focused" and "daydreaming" conditions. However, if clustering (the

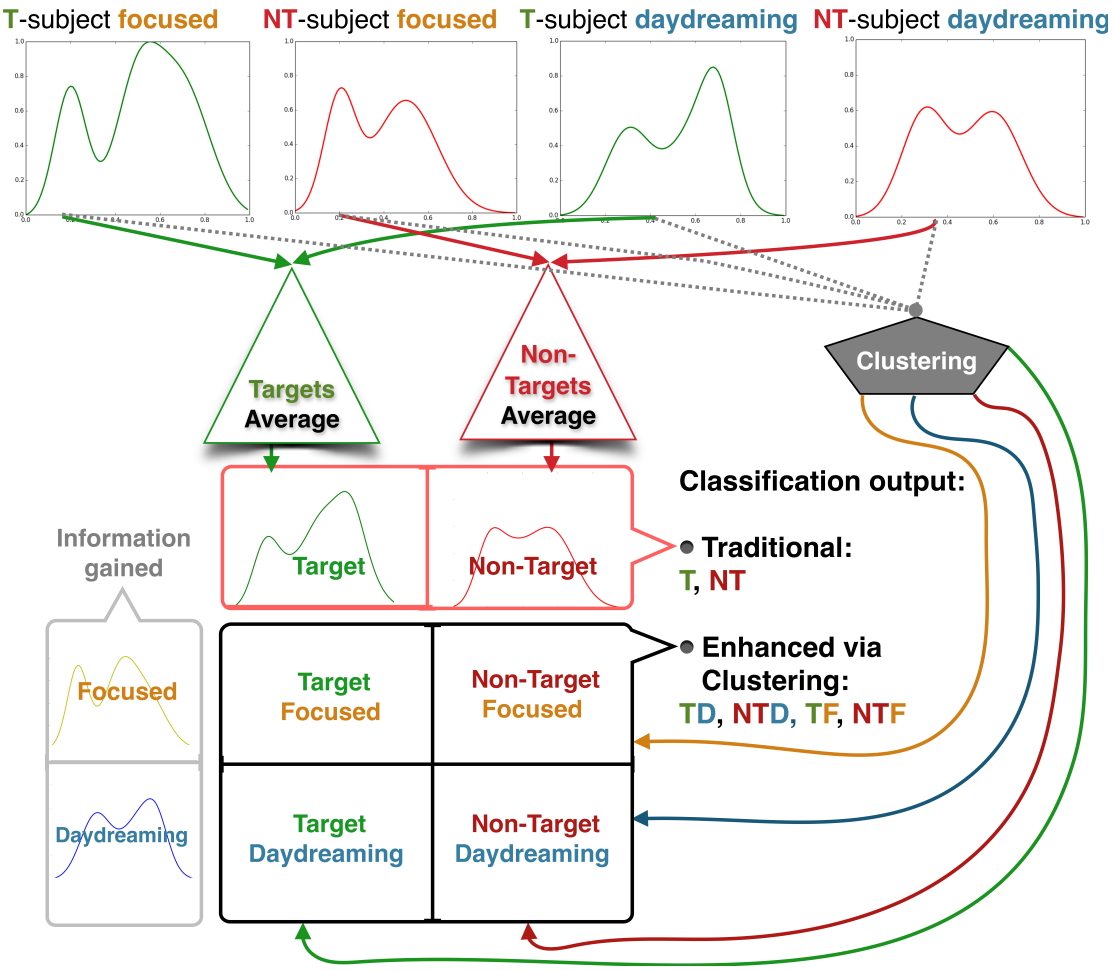

Figure 4.1: Pictorial example of research objectives; enhancing traditional classification via clustering to include the participant's condition.

pentagon) could reveal which ERPs represent a focused mind as opposed to a daydreaming one, not only could we glean information on the wave-forms produced in these states (the two plots to the far left of the figure) but also obtain four new averages representing the response to targets (non-targets) when the user was focused on the task, and corresponding responses when they were daydreaming.

While we recognise that our program is ambitious, it is important to discover to what extent the above can be achieved. In this chapter, we make a start by focusing on studying to what degree a particular form of clustering is suitable for dealing with ERPs. Note that there is a small number of papers where the concept of clustering has been applied on ERPs in topographical analyses ([114, 115]). However, this has been done to differentiate between types of ERP components, e.g. N100, P100, N200 and P300 in terms of location, rather than to group entire responses.

4.1.1 Mental Processes and ERPs

It has been shown that within an experiment, even when the same stimuli are presented, morphologically distinct ERPs can be elicited [96]. Distinguishing between such ERP variants can be important as they may represent different mental processes, yet these ERPs would normally be averaged together with standard averaging techniques [52], thereby blurring any distinctions and limiting our ability to infer mental states and processes from ERPs. Research has shown that conclusively mental processes are also affected by a number of environmental conditions, such as tiredness, food intake, or drugs, which are known to affect the shape, latency and peak values of ERPs [68, 65, 57]. Ultimately, analysing the grand average of ERPs has questionable validity.

This problem also affects BCIs, many of which are based on recognising selected ERPs and transforming them into controls and commands [128, 44, 129, 130]. There are two reasons for this. Firstly, there are still many BCIs that require averaging the responses obtained in multiple repetitions of a stimulus-task pair so as to be able to reliably recognise commands. Since we are essentially unable to differentiate the repetitions where a participant was focused on the task from those where, say, the participant was daydreaming, the signals acquired in the two conditions are inevitably blended, thereby affecting classification accuracy. Secondly, most BCIs rely on forms of machine learning which need to be trained on data with high signal to noise ratio. However, since one cannot distinguish among the alternative mental processes taking place after an event (including some, such as daydreaming, not even triggered by it), we can not build optimum classifiers either.

The literature lacks methods that can successfully classify different kinds of ERPs during an experiment without prior knowledge. The generated ERP could have been classified separately for when the user is distracted, instead of using a classifier trained on data where users were focused. Such a classification process could identify multiple EEG responses, each having a probability of successful acceptance in a separate group measured by TP/FP/TN/FN. By using such a classification method, it is hoped that one could achieve higher levels of communication in BCIs by extending the number of events that the classifier would "recognise" to trigger actions.

To achieve the above, the exploration of several areas is required. In general unsupervised grouping some form of similarity or distance measure is required in order to allocate ERPs to groups. The per-

formance of relevant distance/similarity measures needs to be evaluated; a selection of such measures will be used to describe similarities between groups and the composition of each group of ERPs, i.e. an evaluation of how close groups appear to be (inter connectivity) and how the elements in a group agree with each other (intra-connectivity).

Different similarity methods interpret the data in different ways, so combinations of these methods can also be used to form a more precise measure of similarity. Each group would then be associated with a membership criterion based on the similarity methods used to describe it. With a working grouping criterion the final step would be to consider deriving a way to perform the grouping without referring to stimuli presentation labelling, therefore without knowledge of the stimuli or where and if an ERP was elicited.

4.2 Comparing Similarity Measures in the Presence of White Noise

In order to evaluate and compare methods for clustering ERPs into classes one must have knowledge of the ground truth, or the processes taking place in a participant's mind, for instance in which trials participants were focused on the task and in which they were not. While it is possible to imagine experiments where one could gather real ERP data approximating this, the clarity of such data would be difficult to ascertain with certainty. For this reason, we decided to begin by using synthetic ERP data with characteristics similar to those found in real ERP data, but where we have full control over the ground truth and the signal-to-noise ratio (SNR).

The objective here is to investigate the behaviour of different distance/similarity measures while the SNR is varied to see to what degree the clustering method under study could identify the origin of different ERPs. For this reason, we will evaluate such measures on all possible cluster grouping combinations of such ERPs. Let us call these clusters "partitions". Out of all such partitions, only one, which we will call the "ideal partition", perfectly reproduces the ground truth. All other partitions will have one or more ERP in the wrong class.

4.2.1 Similarity Criteria Based on Centroids

Centroid-based clustering evaluates a distance/similarity measure between the centroids of groups. Here we study and compare three distance/similarity measures: the Euclidean distance, the Manhattan (or city block) distance and finally the Kolmogorov-Smirnoff two-tailed test for similarity of distributions, which is a widely-used non-parametric statistical test.

If p and q are the n-dimensional centroids of two groups, the Euclidean and Manhattan distances are defined as

$$ED(p,q) = \sqrt{\sum_{i=1}^{n}(p_i - q_i)^2} \text{ and } MAN(p,q) = \sum_{i=1}^{n}|p_i - q_i|,$$

respectively. For the Kolmogorov-Smirnov test, we define $KS(p,q)$ as the p-value returned by the KS test when the two data sets $\{q_i\}_{i=1}^{n}$ and $\{p_i\}_{i=1}^{n}$ are fed into it.

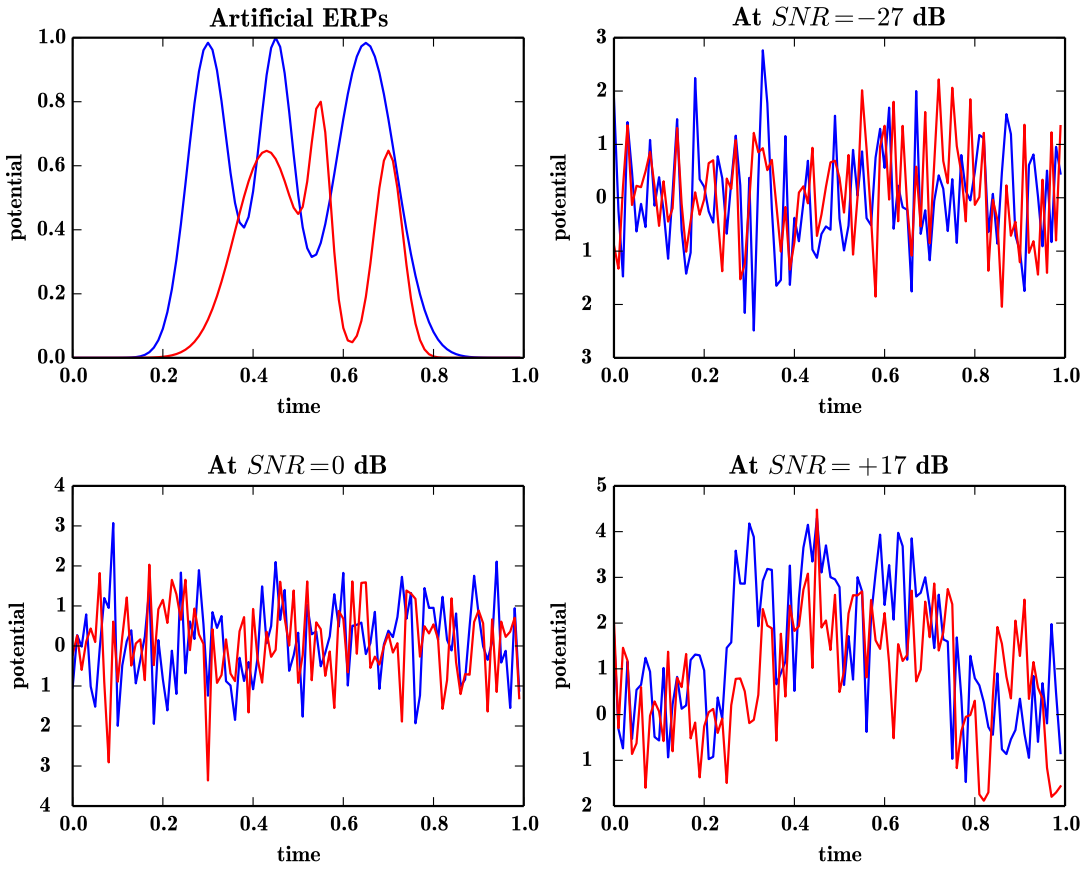

Figure 4.2: On the top-left plot the artificial ERPs used to create the two ERP groups are shown. On the top-right, bottom-right and bottom-left, the same ERPs are degraded by Gaussian noise as described in the text to generate artificial ERPs at different SNR levels.

4.2.2 Artificial ERPs and Noise Generation

Our synthetic ERP set contains 10 ERPs originating from two ideal prototypes obtained by superimposing various mixtures of Gaussians. The prototypes are shown in the top left panel of Figure 4.2.

Noise was introduced in our ERPs by adding zero-mean Gaussian deviates. The SNR was then varied by modifying the standard deviation of these deviates. Examples of the resulting ERPs with different levels of noise degradation are shown in Figure 4.2. Note that SNR=0 dB corresponds to approximately the level of noise of real ERPs, while -27 dB is effectively pure noise, and +17 dB is what one would expect after averaging tens of real ERPs. The two groups of artificial ERPs, shown in the figure, represent two responses elicited by different stimuli.

4.2.3 Error Evaluation

As there are only two classes, partitions can be represented as 10-digit binary strings, where a 0 (1) in a certain position means that the corresponding ERP has been assigned to group 0 (1). Because of this, we can then compare partitions by using the Hamming distance. This is particularly useful when comparing partitions to the ideal partition (that we have conventionally chosen to be string 0000011111

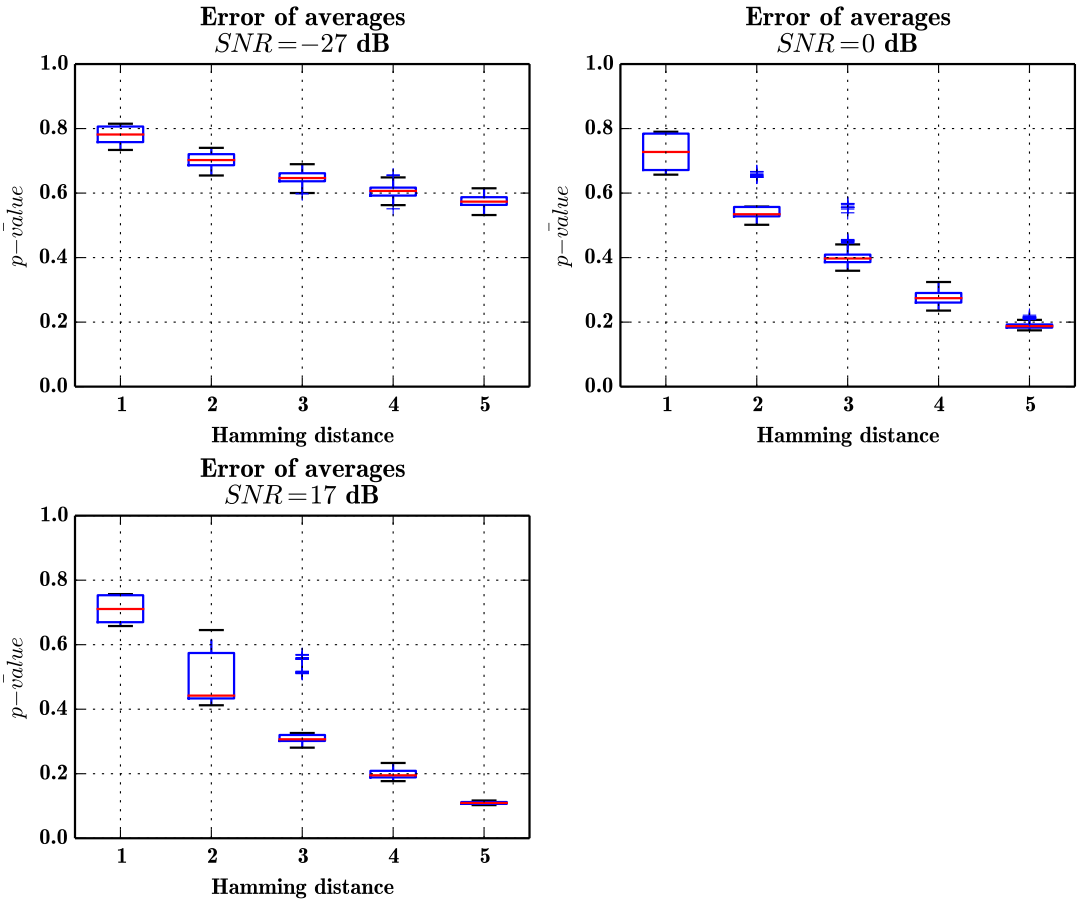

Figure 4.3: Mean p-values of the Wilcoxon rank sum test, comparing the ideal with partitions having non-zero Hamming distances at different SNR levels.

or 1111100000), as this effectively denotes the number of changes needed to transform a partition into the ideal one.

While the Hamming distance is a useful error measure, it is not sufficient to tell us to what degree the assignment of some ERPs to incorrect groups can affect the homogeneity of such groups. This then determines how close measures of central tendency (the mean or the median) for a partition are to those obtained from the ideal partitions.

In order to assess this, for each group in a partition we applied the Wilcoxon rank sum test (the data are paired by time) to see to what degree the actual average resembles the ideal average for the group. Then by averaging the p-values across the groups in the partition, we obtained an indicator of the degree to which clustering errors induce errors in the evaluation of ERP central tendencies.

Figure 4.3, on page 47, reports the mean p-values returned by the test for different partitions at Hamming distance 1, 2, and so on from the ideal partition, and for different SNR levels. As the box plots illustrate, for any one Hamming distance and SNR, there are ample variations in p values, reflecting the fact that assigning some ERPs to the wrong class has very little effect, while with others there can be very serious errors. As was to be expected, we also see that the higher the number of misclassifications the greater the effect on the averages.

4.2.4 Comparison Between the Criteria

The more the centroids of two classes differ (using ED, MAN or KS), the greater the separation between the classes, and the better the clustering. So, we look at the degree to which our similarity measures vary as we vary the Hamming distance of partitions from the optimal partition and the SNR of our synthetic ERPs. The left column of Figures 4.4—4.6 (pages 49—51) reports the $ED(p,q)$, $MAN(p,q)$ and $1-KS(p,q)$ for different values of SNR and different Hamming distances. (We used $1-KS(p,q)$ instead of $KS(p,q)$ so that higher values represent better partitions for all measures.)

If we look at the figures we see that for the highest SNR, the *ED* and *MAN* measures indicate that the best partitions are indeed those with higher *ED* and *MAN* distances, while *KS* is not very sensitive. At SNR=0 dB, a large proportion of sub-optimal partitions are "better" (the centroids of their two groups being more apart) than the ideal partition. At the lowest SNR, the Hamming distance from the ideal partition becomes almost irrelevant in predicting the quality of the clustering.

The probability of a partition other than the ideal appearing to be better than the ideal to a similarity measure is reported in the right column of Figures 4.4—4.6 on pages 49—51. Again, for the highest SNR, we see that the probability of accepting as optimal (the centroids being maximally distant) a configuration that actually is not, decreases with the Hamming distance.

This means that a search algorithm attempting to identify the optimum clusters would probably settle for a configuration that is either the ideal or one with not too many ERPs incorrectly attributed to groups. However, as the SNR decreases, the situation rapidly deteriorates with the exception of *KS* which seems to be more resistant to noise than the other methods.

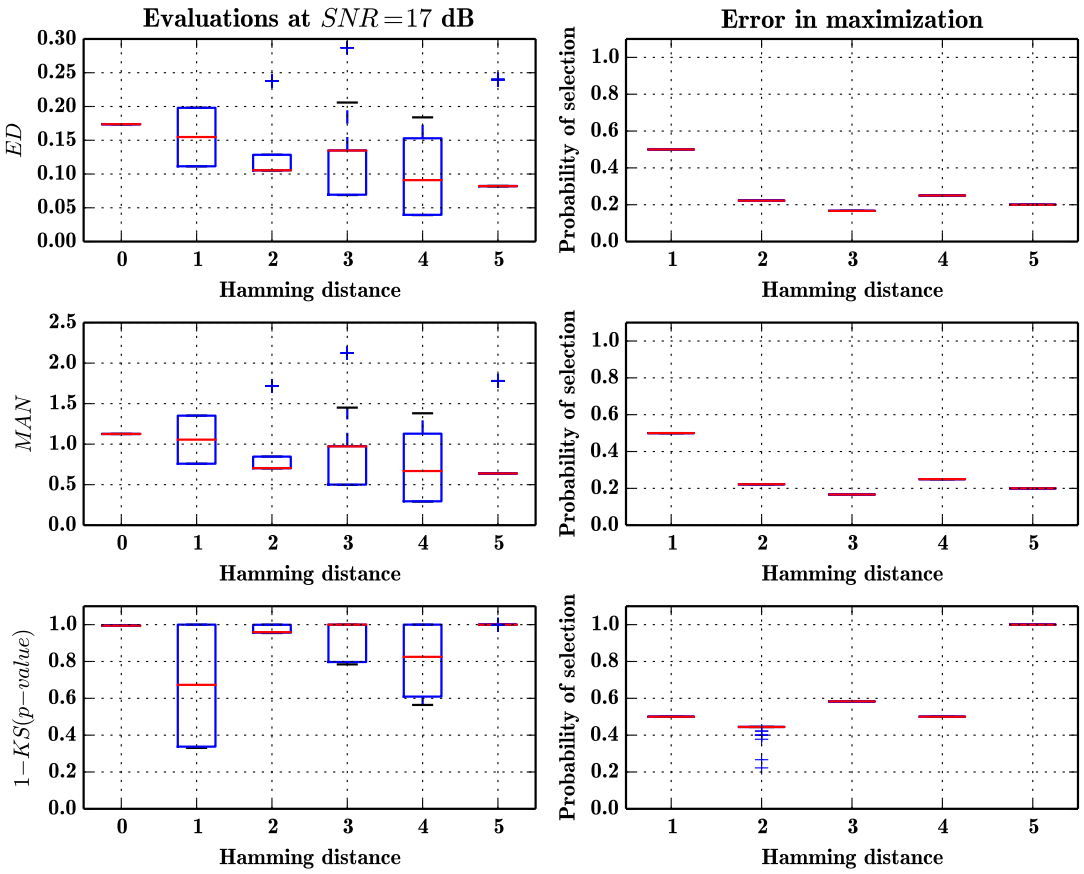

Figure 4.4: Box-plots of the three similarity measures under study (*ED*, *MAN* and *KS*) as a function of the SNR of ERPs and the Hamming distance of partitions (left column), and probability that a partition other than the ideal appears to be better than the ideal to a similarity measure (right column).

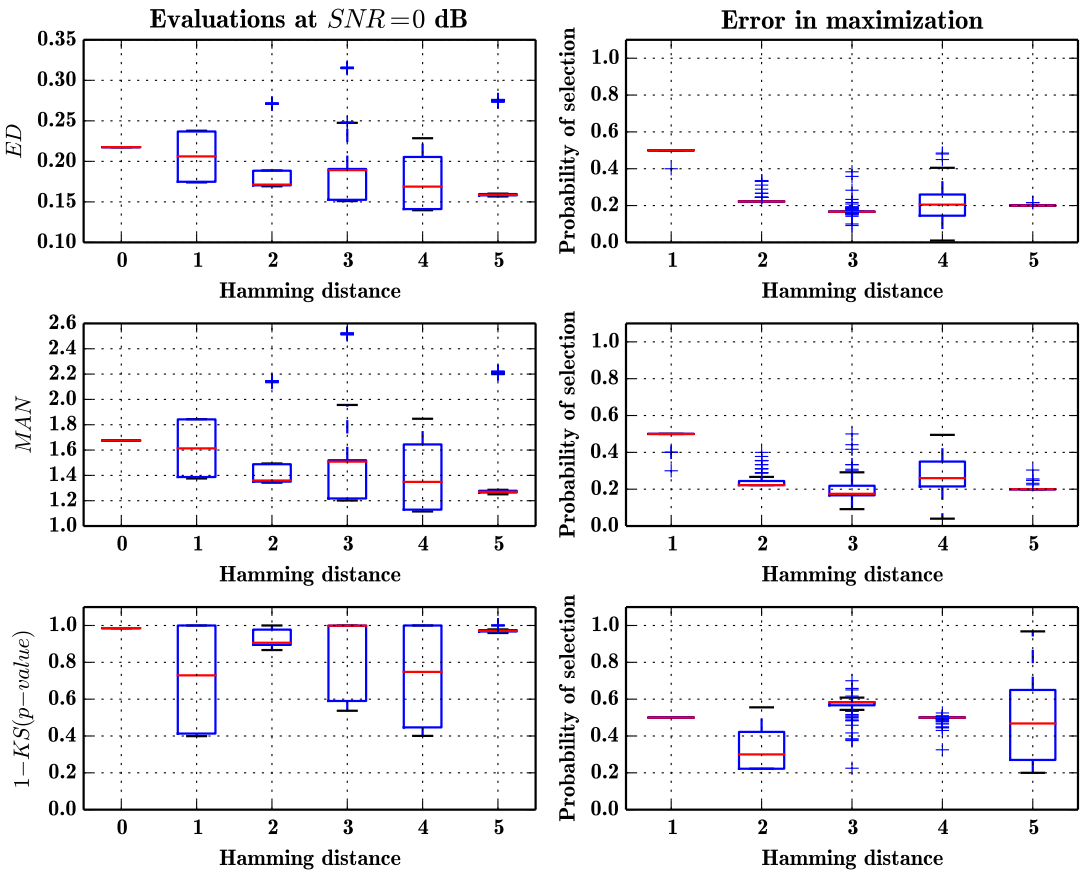

Figure 4.5: Similarly to 4.4

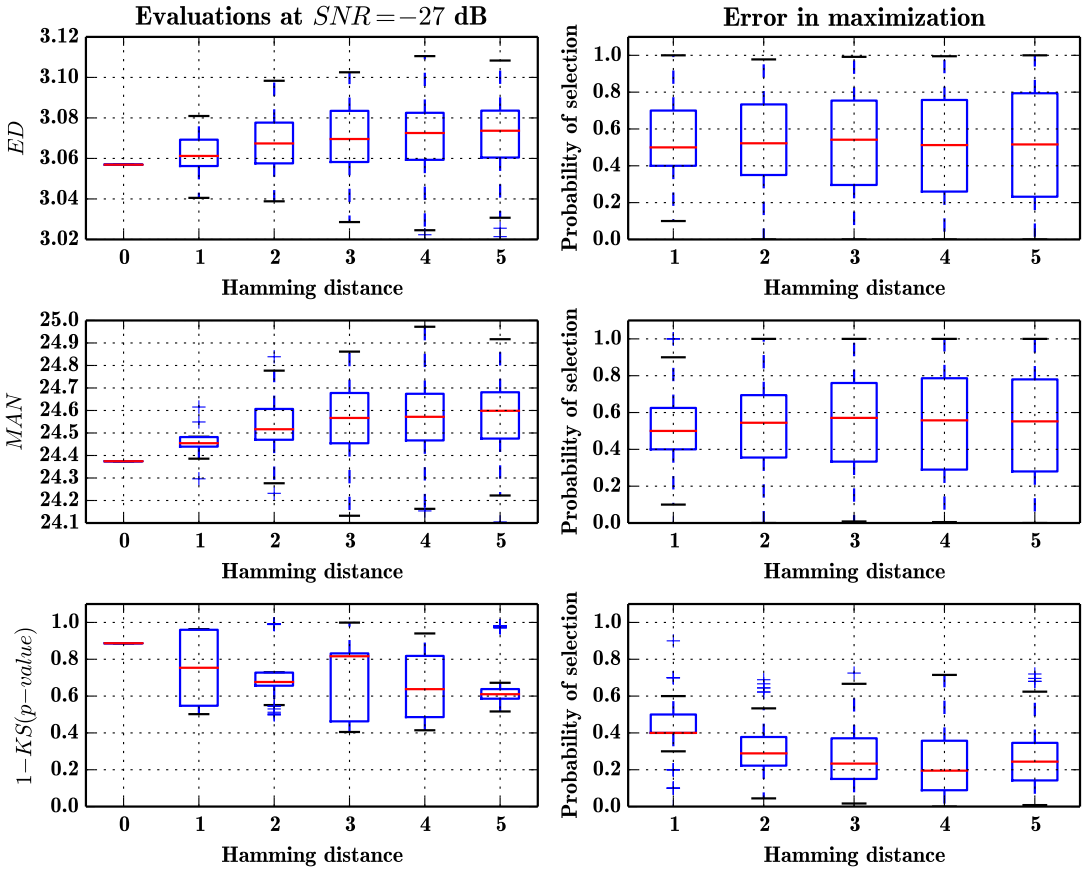

Figure 4.6: Similarly to 4.4

4.3 Extracting Groups of ERPs Based on their Medians

In previous sub-sections, the ERPs were considered to modulate depending on the participants' mental state. Hence clustering was theoretically evaluated in the presence of white noise for two model signals that were similar. In this section, clustering is employed to extract groups of epochs from a set of epochs depending on a confidence level; thus no assumption on the number of different ERPs is made. Although no results from unsupervised methods can be thoroughly explained due to the black box phenomenon, it is possible that they contain valuable information.

The biggest difference between NT and T activity, i.e. the presence of an ERP according to the labelling of epochs during the experiment or not, is the amplitude. In the previous chapter, it was shown that the first derivative can be used as a feature for classification. Differences in magnitudes can be investigated using the Kruskal-Wallis test for similarity in the medians. The following test uses the p-values of the test in order to extract groups of epochs that have similar medians. The precision of the process can be tuned by selecting an appropriate limit for the p-values. Here, the level of significance is chosen to vary from 5% to 1% in steps of 1%. To aid visual comparison the epochs have been passed through a Notch filter prior to the calculation.

In this experiment, the first 5 groups of epochs are extracted. Initially, the algorithm is fed 150 epochs from each class and uses a single electrode location, CPz. At each turn of the process a group of epochs is extracted, while the remaining epochs are retained for the next pass. After 5 passes, if any of the resulting groups contain more than 2 members, then their averages are plotted. Since this is an unsupervised method, the means according to the labels of the experiment are also depicted using dotted lines (see pages 115—130 Figures B.33—B.48 for reference).

Similarly to the exhaustive extraction of groups described above and as shown in Figures B.17—B.32 (pages 99—114), only the first 2 groups are extracted and their averages plotted. The remaining epochs that were not selected during the extraction of these two major groups are averaged and shown. The aim of the process here is to extract epochs containing targets/near-targets and artefacts while rejecting non-targets. As seen in the figures, the unsupervised algorithm manages the above to an extent so that visual interpretation of the results is possible.

As seen in the figures, the results of an almost exhaustive extraction can contain both large and small groups of epochs. It is unknown whether in either the algorithm chose epochs that represent identical or similar processes. In the BCI-Mouse, there were 8 circles used as stimuli, one being the target each time. Since the protocol used was sequential, the expectations of the participant grew during the presentation of the non-target stimuli. If the participant was not focusing and was taken by surprise, a post-target ERP may have been elicited. On the other hand, if the participant was following the rhythm of the experiment in order to respond "as soon as possible" a pre-target ERP could be elicited.

Due to the above, it is not accurate to investigate classes that contain a small number of epochs, since they may even represent similar types of noise. It can be suggested that the first major group that is extracted shows either the group of epochs containing the targets and near targets, or the non-targets and near misses. To make this distinction clearer, the first group of epochs is extracted while the remaining

epochs of the set are used as a second group. Figures B.1—B.16 in pages 99—114, show the results of this evaluation for each participant respectively.

Although the evaluation of these results cannot be conclusive in terms of selection of known ERP, they can be evaluated using the size and the percentage of target epochs selected according to the labels of the experiment. Ideally, for a perfect user, i.e. a participant that elicited distinguishable ERPs only during target stimuli, the number of epochs selected by the algorithm should be 150 and the number of target epochs selected should either be 100% or 0%. In the following Tables 4.1 and 4.2 this evaluation is displayed for all participants. It is clearly visible that for specific confidence levels a higher percentage of Ts, rather than NTs, has been selected. Thus, if one initially *guessed* that the first group represents an *unknown* ERP and that the *unknown* ERP was actually elicited, one has gained knowledge of the shape, latency and amplitude of the *unknown* ERP.

It is evident that the EEG and its first time derivative do not follow the same behaviour at similar grouping strengths. Moreover the resulting percentage of targets selected for a grouping is better with respect to the derivative at 1% when comparing it to the results of the EEG at 5% significance level. V seems to distinguish between epochs even at a low grouping strength, whereas dV requires a smaller significance level to distinguish epochs. This comes as no surprise; the nature of the signals at hand is oscillatory, hence the derivative of the epochs follows a specific rhythm which deviates slightly during target-related epochs. Thus when forming groups using the derivative, by increasing the grouping strength gradually the following can be expected: a) epochs are grouped together depending on the phase of the oscillation; b) artefacts such as eye-blinks and muscle movements are discarded, c) far-target epochs are grouped together, and d) a group contains targets and near-targets.

This method did not aim at reaching the point of having the members allocated into groups as specified by the labels. Its main purpose is to provide researchers with a new tool to investigate responses acquired during an experiment when nothing is known regarding the processes undertaken by the brain of the participant. This is an invaluable addition to the toolbox researchers rely on in EEG analysis, since it may now allow for the extraction of responses that can be meaningful and usable in BCIs without prior knowledge. Although there is no certainty regarding the specifics of the results - stimuli characteristics, property identification or reproducibility - the results of this experiment corroborate the results of the previous chapter that it is better to investigate the derivative of the EEG rather than the EEG itself when evaluating ERPs.

P	5%	4%	3%	2%	1%
1	122(56%)	107(49%)	89(41%)	61(30%)	29(13%)
2	124(61%)	108(52%)	91(47%)	66(35%)	36(18%)
3	119(57%)	107(51%)	87(44%)	56(29%)	26(13%)
4	130(62%)	108(53%)	95(47%)	63(29%)	36(15%)
5	107(51%)	90(45%)	75(39%)	48(25%)	23(13%)
6	120(58%)	98(49%)	78(39%)	62(31%)	36(18%)
7	131(60%)	113(52%)	95(43%)	69(34%)	41(21%)
8	132(68%)	111(58%)	91(51%)	67(39%)	36(21%)
9	131(63%)	108(54%)	93(45%)	70(33%)	36(19%)
10	106(51%)	94(46%)	74(37%)	59(29%)	33(17%)
11	110(55%)	100(51%)	83(42%)	57(33%)	28(17%)
12	141(64%)	122(53%)	106(47%)	75(34%)	42(20%)
13	106(49%)	88(41%)	68(35%)	54(28%)	29(17%)
14	128(57%)	113(51%)	84(39%)	72(33%)	39(19%)
15	144(67%)	129(59%)	102(43%)	74(34%)	45(22%)
16	130(57%)	110(47%)	92(39%)	62(27%)	25(11%)

Table 4.1: Size of extracted group (S) and amount of targets selected (T%) shown as “S(T%)”. Results from using V.

P	5%	4%	3%	2%	1%
1	271(100%)	265(99%)	251(99%)	227(95%)	181(79%)
2	266(100%)	253(100%)	236(98%)	215(91%)	167(78%)
3	271(100%)	263(100%)	255(98%)	223(88%)	161(69%)
4	274(100%)	260(99%)	252(96%)	228(91%)	172(75%)
5	257(99%)	244(97%)	225(92%)	200(85%)	147(65%)
6	259(99%)	253(98%)	242(97%)	223(93%)	168(75%)
7	275(100%)	270(100%)	259(99%)	239(97%)	176(77%)
8	285(100%)	280(100%)	274(100%)	261(100%)	220(94%)
9	267(100%)	255(100%)	243(99%)	212(96%)	156(78%)
10	274(100%)	265(100%)	257(100%)	241(97%)	189(85%)
11	269(100%)	262(100%)	250(99%)	217(92%)	171(79%)
12	256(100%)	253(99%)	242(97%)	219(95%)	163(77%)
13	275(100%)	264(99%)	254(98%)	227(94%)	174(81%)
14	272(100%)	267(99%)	250(97%)	224(90%)	169(75%)
15	268(100%)	263(100%)	249(99%)	232(95%)	179(75%)
16	263(99%)	259(99%)	247(97%)	226(93%)	174(75%)

Table 4.2: Size of extracted group (S) and amount of targets selected (T%) shown as “S(T%)”. Results from using the dV.

4.4 Chapter conclusions

In this thesis I have conducted an initial exploration of the possibility of using unsupervised clustering criteria based on centroids to group ERPs. I found that while the *ED* and *MAN* criteria appear to work well at low noise levels (high SNR), noise influences the quality of the clustering one could hope to obtain very much. Slightly more encouraging results were obtained with the *KS* method.

While the above results are disappointing, there are a number of elements that ought to be considered. Firstly, it should be noted that one of the noise levels used in this exploration (SNR=-27dB) was expected to be problematic, as it is way above that experienced in real ERPs. Secondly, while the analysis was based on generating multiple simulated ERP data sets so as to give more reliable statistics, all such data sets were very small (10 ERPs) compared with the sets used in psycho-physiology or in BCI. Moreover, the artificial ERPs constructed were similar in characteristics, making their comparison very difficult.

The uptake of these results is that even if distance measures are jaded by the increased noise levels, a statistical test could be performed adequately. By using this encouraging outcome, an unsupervised algorithm was implemented using the Kruskal-Wallis test, without using any information from the labels of the experiment, i.e. targets and non-targets. The algorithm was able to extract groups of ERPs which where very similar. Both the EEG and its first-time derivative were able to produce with this method averages that closely represented those derived by the labels of the experiment. The above could eventually result in a very powerful tool for BCIs, enabling the cross-validation of ERP occurrence and separability based solely on the data and not the paradigm specification. Thus, current methods of classification performance could be enhanced.

In future work, there are a number of other techniques to evaluate the quality of partitions, which may well give significantly better results. For instance, it is common to combine measures of separation between clusters with measures of separation within a cluster. A combined criterion such as this would likely produce better ERP clusters. The current status of the research indicates that better results can be obtained. I aim to investigate the above and relative issues in future research.

Chapter 5

Conclusions and Remarks

The future of enhancement-BCIs is intriguing; novel ideas are being developed constantly and the understanding of the working brain is gradually increasing. Currently, the functioning of the brain is considered to be highly related to its anatomy. The communication between different parts and structures of the brain may hold the missing information required to fully integrate man and machine. Ethical and technological limitations regarding the usage of invasive electrodes in humans, for commercial use, hinder the progression of such BCIs.

Enhancement or augmentation of human capabilities using a BCI is surely a topic that will carry forward the next generation of BCIs. Although it was not attempted to create such an algorithm, a major component of this new generation of BCIs has been introduced. In this case clustering was used to reverse the process of the BCI, i.e., instead of using *a priori* information regarding the shape, latency and intensity of an ERP, the BCI reported back to the user these properties. Additionally the usage of a statistical as a measure of similarity in the clustering process can be used to provide confidence at a user-defined level.

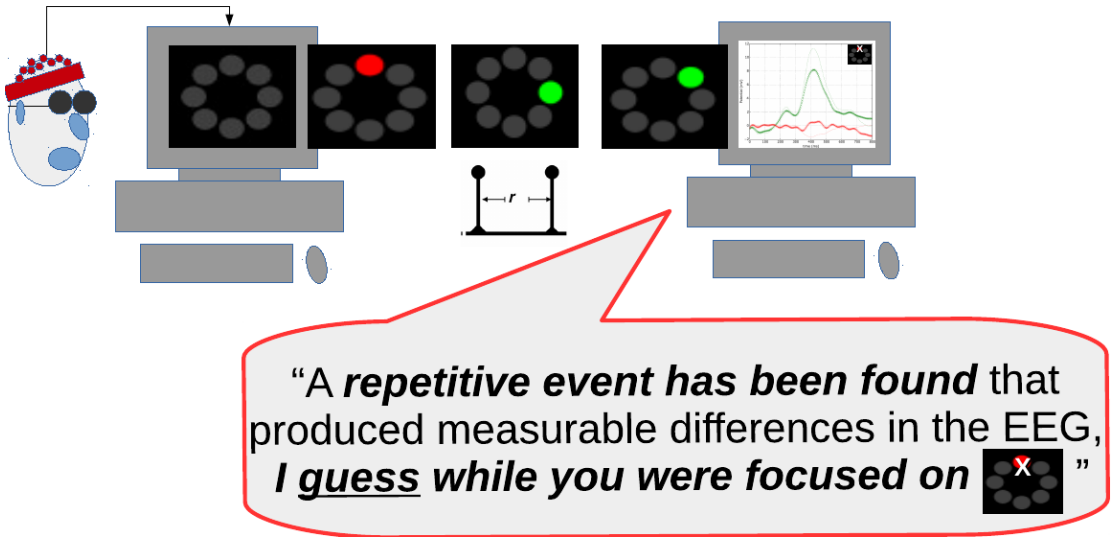

Figure 5.1: Pictorial example of a clustering-enhanced BCI-Mouse.

Although the above conclusions contribute to the area of BCIs by providing a new base for feature extraction and a way to find ERPs without any *a priori* knowledge, these results have further important ramifications. In the case of the BCI-mouse, the clustering algorithm could be used for the machine to *guess* the choice of the user without it being previously specified. As a result the guessing could be used to create a two-way communication link (Figure 5.1) between man and machine.

The above is a revolutionary approach to BCIs, which are currently only recognising ERPs that are known and expected to appear. This current state of BCIs is the major flaw of the DARPA initiative regarding target identification. The initiative contemplated the usage of BCIs for the identification of hostile intentions; satellite images were presented to an analyst while the stimuli used were aeroplanes. In collaboration with NASA JPL, we investigated the application of *brain fusion* for improving the performance of the classification; while the performance was improved, an unknown threat, i.e. other than aeroplanes, could not be detected.

The weakness in the training of machine learning was ultimately responsible for the approach to threat detection being incomplete. The application of the clustering algorithm proposed will enable the BCI to grasp essential properties of ERPs without any training required. Finally and most importantly, the method could also be used to report back to the analyst. Both applications of the proposed algorithm as described above can be used to mitigate the needs of target identification, thus providing an entirely fresh perspective on the topic.

Bibliography

[1] Dimitrios Andreou and Riccardo Poli. Clustering simulated Event-Related Potentials based on similarity of centroids. In *Computer Science and Electronic Engineering Conference (CEEC), 2014 6th*, pages 144–147. IEEE, 2014.

[2] Dimitrios Andreou and Riccardo Poli. Comparing EEG, its time-derivative and their joint use as features in a BCI for 2-D pointer control. In *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference*, pages 5853–5856. IEEE, 2016.

[3] A Stoica, Ana Matran-Fernandez, Dimitrios Andreou, Riccardo Poli, Caterina Cinel, Y Iwashita, and C Padgett. Multi-brain fusion and applications to intelligence analysis. In *SPIE Defense, Security, and Sensing*, pages 87560N–87560N. International Society for Optics and Photonics, 2013.

[4] Anton Nijholt. BCI for games: A 'state of the art' survey. In *Entertainment Computing-ICEC 2008*, pages 225–228. Springer, 2008.

[5] Niels Birbaumer. Breaking the silence: brain–computer interfaces (BCI) for communication and motor control. *Psychophysiology*, 43(6):517–532, 2006.

[6] Joseph N Mak and Jonathan R Wolpaw. Clinical applications of brain-computer interfaces: current state and future prospects. *IEEE reviews in biomedical engineering*, 2:187, 2009.

[7] Lawrence Ashley Farwell and Emanuel Donchin. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and Clinical Neurophysiology*, 70(6):510–523, 1988.

[8] Yongwook Chae, Jaeseung Jeong, and Sungho Jo. Toward brain-actuated humanoid robots: asynchronous direct control using an EEG-based BCI. *Robotics, IEEE Transactions on*, 28(5):1131–1144, 2012.

[9] Iñaki Iturrate, Javier M Antelis, Andrea Kübler, and Javier Minguez. A noninvasive brain-actuated wheelchair based on a P300 neurophysiological protocol and automated navigation. *Robotics, IEEE Transactions on*, 25(3):614–627, 2009.

[10] Álvaro Barbero and Moritz Grosse-Wentrup. Biased feedback in brain-computer interfaces. *Journal of neuroengineering and rehabilitation*, 7(1):1, 2010.

[11] R Ortner, DC Irimia, J Scharinger, and C Guger. A motor imagery based brain-computer interface for stroke rehabilitation. *Stud Health Technol Inform*, 181:319–23, 2012.

[12] Bo Hjorth. EEG analysis based on time domain properties. *Electroencephalography and clinical neurophysiology*, 29(3):306–310, 1970.

[13] Carmen Vidaurre, Nicole Krämer, Benjamin Blankertz, and Alois Schlögl. Time domain parameters as a feature for EEG-based brain–computer interfaces. *Neural Networks*, 22(9):1313–1319, 2009.

[14] B.R. Greene, S. Faul, W.P. Marnane, G. Lightbody, I. Korotchikova, and G.B. Boylan. A comparison of quantitative EEG features for neonatal seizure detection. *Clinical Neurophysiology*, 119(6):1248 – 1261, 2008.

[15] Yohei Tomita, Francois-Benoit Vialatte, Gérard Dreyfus, Yasue Mitsukura, Hovagim Bakardjian, and Andrzej Cichocki. Bimodal BCI using simultaneously NIRS and EEG. *Biomedical Engineering, IEEE Transactions on*, 61(4):1274–1284, 2014.

[16] Cuntai Guan, Manoj Thulasidas, and Jiankang Wu. High performance P300 speller for brain-computer interface. In *Biomedical Circuits and Systems, 2004 IEEE International Workshop on*, pages S3–5. IEEE, 2004.

[17] Manoj Thulasidas, Cuntai Guan, and Jiankang Wu. Robust classification of EEG signal for brain-computer interface. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 14(1):24, 2006.

[18] Bang hua Yang, Guo zheng Yan, Rong guo Yan, and Ting Wu. Adaptive subject-based feature extraction in brain-computer interfaces using wavelet packet best basis decomposition. *Medical Engineering & Physics*, 29(1):48 – 53, 2007.

[19] WEI-YEN HSU. Continuous EEG signal analysis for asynchronous BCI application. *International Journal of Neural Systems*, 21(04):335–350, 2011. PMID: 21809479.

[20] Wu Ting, Yan Guo-zheng, Yang Bang-hua, and Sun Hong. EEG feature extraction based on wavelet packet decomposition for brain computer interface. *Measurement*, 41(6):618 – 625, 2008.

[21] Bang hua Yang, Guo zheng Yan, Ting Wu, and Rong guo Yan. Subject-based feature extraction using fuzzy wavelet packet in brain-computer interfaces. *Signal Processing*, 87(7):1569 – 1574, 2007.

[22] V. Bostanov. BCI competition 2003-data sets Ib and IIb: feature extraction from event-related brain potentials with the continuous wavelet transform and the t-value scalogram. *Biomedical Engineering, IEEE Transactions on*, 51(6):1057–1061, June 2004.

[23] O. Carrera-Leon, J.M. Ramirez, V. Alarcon-Aquino, M. Baker, D. D'Croz-Baron, and P. Gomez-Gil. A motor imagery BCI experiment using wavelet analysis and spatial patterns feature extraction. In *Engineering Applications (WEA), 2012 Workshop on*, pages 1–6, May 2012.

[24] Qi Xu, Hui Zhou, Yongji Wang, and Jian Huang. Fuzzy support vector machine for classification of EEG signals using wavelet-based features. *Medical Engineering & Physics*, 31(7):858 – 865, 2009.

[25] Carles Escera, Kimmo Alho, Erich Schröger, and István Winkler. Involuntary attention and distractibility as evaluated with event-related brain potentials. *Audiology and Neurotology*, 5(3-4):151–166, 2000.

[26] Marc Jeannerod. Mental imagery in the motor context. *Neuropsychologia*, 33(11):1419–1432, 1995.

[27] David Regan. Steady-state evoked potentials. *JOSA*, 67(11):1475–1489, 1977.

[28] Gyorgy Buzsaki. *Rhythms of the Brain*. Oxford University Press, 2006.

[29] E Basar, C Basar-Eroglu, S Karakas, and M Schurmann. Brain oscillations in perception and memory. *International Journal of Psychophysiology*, 35(2-3):95 – 124, 2000.

[30] J. L. R. Rubenstein and M. M. Merzenich. Model of autism: increased ratio of excitation/inhibition in key neural systems. *Genes, Brain and Behavior*, 2(5):255–267, 2003.

[31] Wolfgang Klimesch. Alpha-band oscillations, attention, and controlled access to stored information. *Trends in cognitive sciences*, 2012.

[32] Christoph von der Malsburg. Binding in models of perception and brain function. *Current Opinion in Neurobiology*, 5(4):520–526, 1995.

[33] GN Borisyuk, RM Borisyuk, Yakov Borisovich Kazanovich, and Genrikh Romanovich Ivanitskii. Models of neural dynamics in brain information processing-the developments of'the decade'. *Physics-Uspekhi*, 45(10):1073–1095, 2002.

[34] Isabelle Brocas and Juan Carrillo. Reason, emotion and information processing in the brain. *NONE – UPDATE*, 2008.

[35] Verner J. Knott and Peter H. Venables. EEG alpha correlates of non-smokers, smokers, smoking, and smoking deprivation. *Psychophysiology*, 14(2):150–156, 1977.

[36] T.F. Oostendorp, J. Delbeke, and Dick F. Stegeman. The conductivity of the human skull: results of in vivo and in vitro measurements. *Biomedical Engineering, IEEE Transactions on*, 47(11):1487–1492, Nov 2000.

[37] Jorma O Ollikainen, Marko Vauhkonen, Pasi A Karjalainen, and Jari P Kaipio. Effects of local skull inhomogeneities on EEG source estimation. *Medical Engineering & Physics*, 21(3):143 – 154, 1999.

[38] Francesco Brigo, Rosario Cicero, Antonio Fiaschi, and Luigi Giuseppe Bongiovanni. The breach rhythm. *Clinical Neurophysiology*, 122(11):2116 – 2120, 2011.

[39] T.C. Ferree, K. J. Eriksen, and D.M. Tucker. Regional head tissue conductivity estimation for improved EEG analysis. *Biomedical Engineering, IEEE Transactions on*, 47(12):1584–1592, Dec 2000.

[40] Katrina Elizabeth Wendel. The influence of tissue conductivity and head geometry on EEG measurement sensitivity distributions. *Tampereen teknillinen yliopisto. Julkaisu-Tampere University of Technology. Publication; 900*, 2010.

[41] P Bušek and D Kemlink. The influence of the respiratory cycle on the EEG. *Physiological research*, 54:327–33, 2005.

[42] Enoch Callaway and Robert S. Layne. Interaction between the visual evoked response and two spontaneous biological rhythms: the EEG alpha cycle and the cardiac arousal cycle. *Annals of the New York Academy of Sciences*, 112(1):421–431, 1964.

[43] Setare Amiri, Reza Fazel-Rezai, and Vahid Asadpour. A review of hybrid brain-computer interface systems. *Adv. in Hum.-Comp. Int.*, 2013:1:1–1:1, January 2013.

[44] E. Donchin and Yael Arbel. P300 Based Brain Computer Interfaces: A Progress Report. *Augmented Cognition*, 2009. Speed-Accuracy Tradeoffs.

[45] Mehrdad Fatourechi, Ali Bashashati, Rabab K. Ward, and Gary E. Birch. EMG and EOG artifacts in brain computer interface systems: A survey. *Clinical Neurophysiology*, 118(3):480 – 494, 2007.

[46] Emanuel Donchin, Walter Ritter, W Cheyne McCallum, et al. Cognitive psychophysiology: The endogenous components of the erp. *Event-related brain potentials in man*, pages 349–411, 1978.

[47] Gregory McCarthy and Emanuel Donchin. A metric for thought: a comparison of P300 latency and reaction time. *Science*, 211(4477):77–80, 1981.

[48] Robert D Sidman, Martin R Ford, Gloria Ramsey, and Christine Schlichting. Age-related features of the resting and P300 auditory evoked responses using the dipole localization method and cortical imaging technique. *Journal of neuroscience methods*, 33(1):23–32, 1990.

[49] Mark Polak and Aleksandar Kostov. Feature extraction in development of brain-computer interface: a case study. In *Engineering in Medicine and Biology Society, 1998. Proceedings of the 20th Annual International Conference of the IEEE*, volume 4, pages 2058–2061. IEEE, 1998.

[50] José del R Millán, Marco Franzé, Josep Mouriño, Febo Cincotti, and Fabio Babiloni. Relevant EEG features for the classification of spontaneous motor-related tasks. *Biological cybernetics*, 86(2):89–95, 2002.

[51] Fabien Lotte, Marco Congedo, Anatole Lécuyer, Fabrice Lamarche, and Bruno Arnaldi. A review of classification algorithms for EEG-based brain–computer interfaces. *Journal of neural engineering*, 4(2):R1, 2007.

[52] Steven J Luck. *An introduction to the event-related potential technique (cognitive neuroscience)*. The MIT Press, 2005.

[53] Shravani Sur and VK Sinha. Event-related potential: An overview. *Industrial psychiatry journal*, 18(1):70, 2009.

[54] Robert M Chapman and Henry Bragdon. Evoked responses to numerical and non-numerical visual stimuli while problem solving. *Nature*, 1964.

[55] Samuel Sutton, Margery Braren, Joseph Zubin, and E. R. John. Evoked-potential correlates of stimulus uncertainty. *Science*, 150(3700):1187–1188, 1965.

[56] Samuel Sutton, Patricia Tueting, Joseph Zubin, and E. R. John. Information delivery and the sensory evoked potential. *Science*, 155(3768):1436–1439, 1967.

[57] Craig J Gonsalvez and John Polich. P300 amplitude is determined by target-to-target interval. *Psychophysiology*, 39(3):388–396, 2002.

[58] J Peter Rosenfeld, Julianne R Biroschak, Melissa J Kleschen, and Kyle M Smith. Subjective and objective probability effects on P300 amplitude revisited. *Psychophysiology*, 42(3):356–359, 2005.

[59] Carlos M Gomez, Angelica Flores, Marcia R Digiacomo, Alfredo Ledesma, and Javier Gonzalez-Rosa. P3a and P3b components associated to the neurocognitive evaluation of invalidly cued targets. *Neuroscience letters*, 430(2):181–185, 2008.

[60] John Polich. Updating P300: an integrative theory of P3a and P3b. *Clinical Neurophysiology*, 118(10):2128, 2007.

[61] Marco D Comerchero and John Polich. P3a, perceptual distinctiveness, and stimulus modality. *Cognitive Brain Research*, 7(1):41–48, 1998.

[62] Tom Eichele, Karsten Specht, Matthias Moosmann, Marijtje LA Jongsma, Rodrigo Quian Quiroga, Helge Nordby, and Kenneth Hugdahl. Assessing the spatiotemporal evolution of neuronal activation with single-trial event-related potentials and functional MRI. *Proceedings of the National Academy of Sciences of the United States of America*, 102(49):17798–17803, 2005.

[63] W Ritter and D S Ruchkin. A review of event-related potential components discovered in the context of studying P3. *Ann. N. Y. Acad. Sci.*, 658:1–32, July 1992.

[64] Daran Ravden and John Polich. On P300 measurement stability: habituation, intra-trial block variation, and ultradian rhythms. *Biological Psychology*, 51(1):59–76, 1999.

[65] Heather M Gray, Nalini Ambady, William T Lowenthal, and Patricia Deldin. P300 as an index of attention to self-relevant stimuli. *Experimental Social Psychology*, 40(2):216–224, 2004.

[66] J Polich. On the relationship between EEG and P300: individual differences, aging, and ultradian rhythms. *International journal of psychophysiology*, 26(1):299–317, 1997.

[67] Daran Ravden and John Polich. On P300 measurement stability: habituation, intra-trial block variation, and ultradian rhythms. *Biological psychology*, 51(1):59–76, 1999.

[68] Sven P Heinrich, David Marhöfer, and Michael Bach. Cognitive visual acuity estimation based on the event-related potential P300 component. *Clinical Neurophysiology*, 121(9):1464–1472, 2010.

[69] CEM Van Beijsterveldt and GCM Van Baal. Twin and family studies of the human electroencephalogram: a review and a meta-analysis. *Clinical Neurophysiology*, 61(1):111–138, 2002.

[70] Kimron L Shapiro. The attentional blink: The brain's" eyeblink". *Current Directions in Psychological Science*, 3(3):86–89, 1994.

[71] Anna Rita Bentivoglio, Susan B Bressman, Emanuele Cassetta, Donatella Carretta, Pietro Tonali, and Alberto Albanese. Analysis of blink rate patterns in normal subjects. *Movement Disorders*, 12(6):1028–1034, 1997.

[72] Philipp P Caffier, Udo Erdmann, and Peter Ullsperger. Experimental evaluation of eye-blink parameters as a drowsiness measure. *European journal of applied physiology*, 89(3-4):319–325, 2003.

[73] Timo Stein, Ignacio Vallines, and Werner X Schneider. Primary visual cortex reflects behavioral performance in the attentional blink. *Neuroreport*, 19(13):1277–1281, 2008.

[74] Ulrich Hoffmann, Ashkan Yazdani, Jean-Marc Vesin, and Touradj Ebrahimi. Bayesian Feature Selection Applied In a P300 Brain- Computer Interface. In *proceedings of 16th European Signal Processing Conference (EUSIPCO 2008)*, Lausanne , Switzerland, 2008.

[75] Jaime F Delgado Saa and Miguel Sotaquirá Gutierrez. EEG signal classification using power spectral features and linear discriminant analysis: A brain computer interface application. In *Eighth Latin American and Caribbean Conference for Engineering and Technology*, pages 1–7, 2010.

[76] Dennis J McFarland, Charles W Anderson, K Muller, Alois Schlogl, and Dean J Krusienski. BCI meeting 2005-workshop on BCI signal processing: feature extraction and translation. *IEEE transactions on neural systems and rehabilitation engineering*, 14(2):135, 2006.

[77] C. Guerrero-Mosquera, M. Verleysen, and A.N. Vazquez. EEG feature selection using mutual information and support vector machine: A comparative analysis. In *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, pages 4946–4949, Aug 2010.

[78] Kai Keng Ang, Zheng Yang Chin, Haihong Zhang, and Cuntai Guan. Mutual information-based selection of optimal spatial-temporal patterns for single-trial EEG-based BCIs. *Pattern Recognition*, 45(6):2137 – 2144, 2012. Brain Decoding.

[79] Irena Koprinska. Feature selection for brain-computer interfaces. In Thanaruk Theeramunkong, Cholwich Nattee, PauloJ.L. Adeodato, Nitesh Chawla, Peter Christen, Philippe Lenca, Josiah Poon, and Graham Williams, editors, *New Frontiers in Applied Data Mining*, volume 5669 of *Lecture Notes in Computer Science*, pages 106–117. Springer Berlin Heidelberg, 2010.

[80] C.W. Anderson and M.J. Kirby. EEG subspace representations and feature selection for brain-computer interfaces. In *Computer Vision and Pattern Recognition Workshop, 2003. CVPRW '03. Conference on*, volume 5, pages 51–51, June 2003.

[81] E. Yom-Tov and G.F. Inbar. Feature selection for the classification of movements from single movement-related potentials. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 10(3):170–177, Sept 2002.

[82] N.S. Dias, M. Kamrunnahar, P.M. Mendes, S.J. Schiff, and J.H. Correia. Feature selection on movement imagery discrimination and attention detection. *Medical & Biological Engineering & Computing*, 48(4):331–341, 2010.

[83] T.H. Dat and Cuntai Guan. Feature selection based on Fisher ratio and mutual information analyses for robust brain computer interface. In *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, volume 1, pages I–337–I–340, April 2007.

[84] Jinyi Long, Zhenghui Gu, Yuanqing Li, Tianyou Yu, Feng Li, and Ming Fu. Semi-supervised joint spatio-temporal feature selection for P300-based BCI speller. *Cognitive Neurodynamics*, 5(4):387–398, 2011.

[85] D. Flotzinger, M. Pregenzer, and G. Pfurtscheller. Feature selection with distinction sensitive learning vector quantisation and genetic algorithms. In *Neural Networks, 1994. IEEE World Congress on Computational Intelligence., 1994 IEEE International Conference on*, volume 6, pages 3448–3451 vol.6, Jun 1994.

[86] Zenglin Xu, Rong Jin, Jieping Ye, Michael R. Lyu, and Irwin King. Non-monotonic feature selection. In *Proceedings of the 26th Annual International Conference on Machine Learning*, ICML '09, pages 1145–1152, New York, NY, USA, 2009. ACM.

[87] Yijun Wang, Shangkai Gao, and Xiaorong Gao. Common spatial pattern method for channel selelction in motor imagery based brain-computer interface. In *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*, pages 5392–5395, Jan 2005.

[88] G. Dornhege, B. Blankertz, G. Curio, and K. Muller. Boosting bit rates in noninvasive EEG single-trial classifications by feature combination and multiclass paradigms. *Biomedical Engineering, IEEE Transactions on*, 51(6):993–1002, June 2004.

[89] Guido Dornhege, Benjamin Blankertz, Gabriel Curio, and Klaus-Robert Müller. Combining features for BCI. In *Advances in Neural Information Processing Systems*, pages 1115–1122, 2002.

[90] CÃ©line Bugli and Philippe Lambert. Comparison between principal component analysis and independent component analysis in electroencephalograms modelling. *Biometrical journal*, 49(2):312–327, 2007.

[91] Hyunjin Yoon, Kiyoung Yang, and C. Shahabi. Feature subset selection and feature ranking for multivariate time series. *Knowledge and Data Engineering, IEEE Transactions on*, 17(9):1186–1198, Sept 2005.

[92] Tzyy-Ping Jung, Scott Makeig, Colin Humphries, Te-Won Lee, Martin J Mckeown, Vicente Iragui, and Terrence J Sejnowski. Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, 37(02):163–178, 2000.

[93] Laurent Albera, Amar Kachenoura, Pierre Comon, Ahmad Karfoul, Fabrice Wendling, Lotfi Senhadji, and Isabelle Merlet. ICA-based EEG denoising: a comparative analysis of fifteen methods. *Bulletin of the Polish Academy of Sciences*, 60, October 2012.

[94] David A. Peterson, James N. Knight, Michael J. Kirby, Charles W. Anderson, and Michael H. Thaut. Feature selection and blind source separation in an EEG-based brain-computer interface. *EURASIP J. Appl. Signal Process.*, 2005:3128–3140, January 2005.

[95] Riccardo Poli, Luca Citi, Mathew Salvaris, Caterina Cinel, and Francisco Sepulveda. Eigenbrains: The free vibrational modes of the brain as a new representation for EEG. In *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, pages 6011–6014. IEEE, 2010.

[96] R. Poli, C. Cinel, L. Citi, and F. Sepulveda. Reaction-time binning: A simple method for increasing the resolving power of ERP averages. *Psychophysiology, 47, pg. 467-485*, 2010. Stimulus-locked Averaging, Response-locked Averaging, ERP-locked averaging, Model-based ERP Reconstruction, Grand Averages and Averages Across Subjects, Response-time binning.

[97] Stefan Debener, Scott Makeig, Arnaud Delorme, and Andreas K Engel. What is novel in the novelty oddball paradigm? Functional significance of the novelty P3 event-related potential as revealed by independent component analysis. *Cognitive Brain Research*, 22(3):309–321, 2005.

[98] AndrÃ© Beauducel, Stefan Debener, et al. Misallocation of variance in event-related potentials: simulation studies on the effects of test power, topography, and baseline-to-peak versus principal component quantifications. *Neuroscience Methods*, 124(1):103, 2003.

[99] Leslie D Thede. *Practical analog and digital filter design*. Artech House New Jersey, 2005.

[100] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller. Optimal spatial filtering of single trial EEG during imagined hand movement. *Rehabilitation Engineering, IEEE Transactions on*, 8(4):441–446, Dec 2000.

[101] G. Garcia-Molina and Danhua Zhu. Optimal spatial filtering for the steady state visual evoked potential: BCI application. In *Neural Engineering (NER), 2011 5th International IEEE/EMBS Conference on*, pages 156–160, April 2011.

[102] Gabriel Pires, Urbano Nunes, and Miguel Castelo-Branco. Statistical spatial filtering for a P300-based BCI: Tests in able-bodied, and patients with cerebral palsy and amyotrophic lateral sclerosis. *Journal of Neuroscience Methods*, 195(2):270 – 281, 2011.

[103] John Canny. A computational approach to edge detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, PAMI-8(6):679–698, Nov 1986.

[104] Bob Zhang, Lin Zhang, Lei Zhang, and Fakhri Karray. Retinal vessel extraction by matched filter with first-order derivative of gaussian. *Computers in biology and medicine*, 40(4):438–445, 2010.

[105] Riccardo Poli and Guido Valli. An algorithm for real-time vessel enhancement and detection. *Computer methods and programs in Biomedicine*, 52(1):1–22, 1997.

[106] Paul S Addison. *The illustrated wavelet transform handbook: introductory theory and applications in science, engineering, medicine and finance*. CRC press, 2002.

[107] Christopher E Heil and David F Walnut. Continuous and discrete wavelet transforms. *SIAM review*, 31(4):628–666, 1989.

[108] E. Donchin. Averaged evoked potentials and uncertainty resolution. *Psychon. Sci. Vol. 12 (3)*, 1968. Study of averaged evoked potentials to near-threshold stimuli - Confirmation of Sutton's study(1967).

[109] B. Lütkenhöner, M. Hoke, and Ch. Pantev. Possibilities and limitations of weighted averaging. *Biological Cybernetics*, 52(6):409–416, 1985.

[110] Mark M. Stecker. Generalized averaging and noise levels in evoked responses. *Computers in Biology and Medicine*, 247-265(30), July 2000.

[111] Z. Leonowicz, J. Karvanen, and S. L. Shishkin. Trimmed estimators for robust averaging of event-related potentials. *Neuroscience Methods*, 2004.

[112] E. Bataillou, E. Thierry, H. Rix, and O. Meste. Weighted averaging using adaptive estimation of the weights. *Signal Processing*, 44(1):51–66, 1995.

[113] Sivylla E. Paraskevopoulou, Di Wu, Amir Eftekhar, and Timothy G. Constandinou. Hierarchical adaptive means (ham) clustering for hardware-efficient, unsupervised and real-time spike sorting. *Journal of Neuroscience Methods*, 235:145 – 156, 2014.

[114] MicahM. Murray, Denis Brunet, and ChristophM. Michel. Topographic ERP analyses: A step-by-step tutorial review. *Brain Topography*, 20(4):249–264, 2008.

[115] Dejing Dou, Gwen Frishkoff, Jiawei Rong, Robert Frank, Allen Malony, and Don Tucker. Development of neuroelectromagnetic ontologies (nemo): a framework for mining brainwave ontologies. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 270–279. ACM, 2007.

[116] S. Aviyente, L.A.W. Brakel, R.K. Kushwaha, M. Snodgrass, Howard Shevrin, and W.J. Williams. Characterization of event related potentials using information theoretic distance measures. *Biomedical Engineering, IEEE Transactions on*, 51:737–743, 2004.

[117] Selin Aviyente. Information-theoretic signal processing on the time-frequency plane and applications. In *Signal Processing Conference, 2005 13th European*, pages 1–4. IEEE, 2005.

[118] R Quian Quiroga and H Garcia. Single-trial event-related potentials with wavelet denoising. *Clinical Neurophysiology*, 114(2):376–390, 2003.

[119] M.Sasha John, Andrew Dimitrijevic, and Terence W Picton. Weighted averaging of steady-state responses. *Clinical Neurophysiology*, 112(3):555–562, 2001.

[120] Scott Makeig. Auditory event-related dynamics of the EEG spectrum and effects of exposure to tones. *Electroencephalography and Clinical Neurophysiology*, 86(4):283–293, 1993.

[121] Anthony T Cacace and Dennis J McFarland. Spectral dynamics of electroencephalographic activity during auditory information processing. *Hearing Research*, 176(1-2):25–41, 2003.

[122] Alexander V Kramarenko and Uner Tan. Brief communication validity of spectral analysis of evoked potentials in brain research. *International journal of neuroscience*, 112(4):489–499, 2002.

[123] Luca Citi, Riccardo Poli, Caterina Cinel, and Francisco Sepulveda. P300-based BCI mouse with genetically-optimized analogue control. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 16(1):51–61, 2008.

[124] Riccardo Poli, Mathew Salvaris, and Caterina Cinel. Evolution of a brain-computer interface mouse via genetic programming. In *Genetic Programming*, pages 203–214. Springer, 2011.

[125] M. Salvaris, C. Cinel, R. Poli, L. Citi, and F. Sepulveda. Exploring multiple protocols for a brain-computer interface mouse. In *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, pages 4189–4192, Aug 2010.

[126] Mathew Salvaris, Caterina Cinel, Luca Citi, and Riccardo Poli. Novel protocols for P300-based brain–computer interfaces. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 20(1):8–17, 2012.

[127] PM Quilter, BB MacGillivray, and DG Wadbrook. The removal of eye movement artefact from EEG signals using correlation techniques. In *Random Signal Analysis, IEEE Conference Publication*, volume 159, pages 93–100, 1977.

[128] Jonathan R Wolpaw, et. al. Brain-computer interfaces for communication and control. *Clinical Neurophysiology*, 113(6):767–791, 2002.

[129] S. Noshadi and S. Es'haghi. Basic Information about BCI Systems. *Research Journal of Applied Sciences, Engineering and Technology*, 5(11):3144–3151, April 2013.

[130] Mandeep Kaur, P. Ahmed, and M. Qasim Rafiq. Analyzing EEG based Neurological Phenomenon. *International Journal of Computer Applications*, 57(17), November 2012.

Appendices

Appendix A

Derivatives

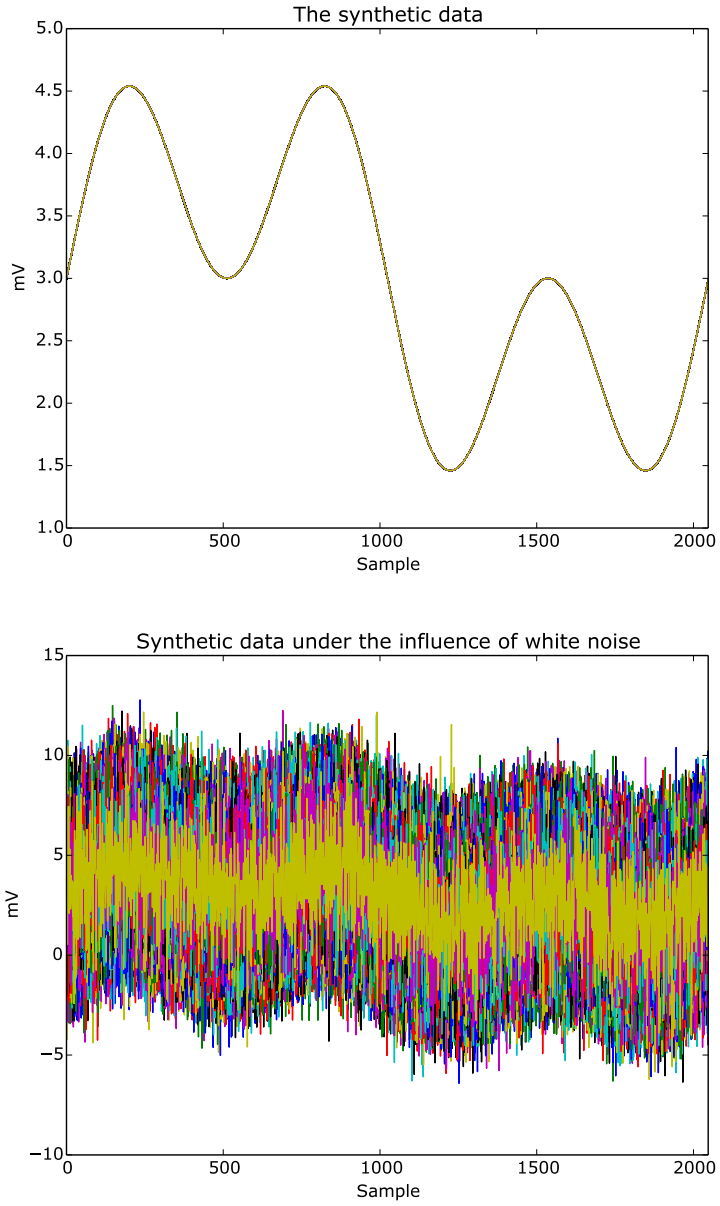

Figure A.1: (a) A synthetic ERP, (b) resulting data under white noise influence.

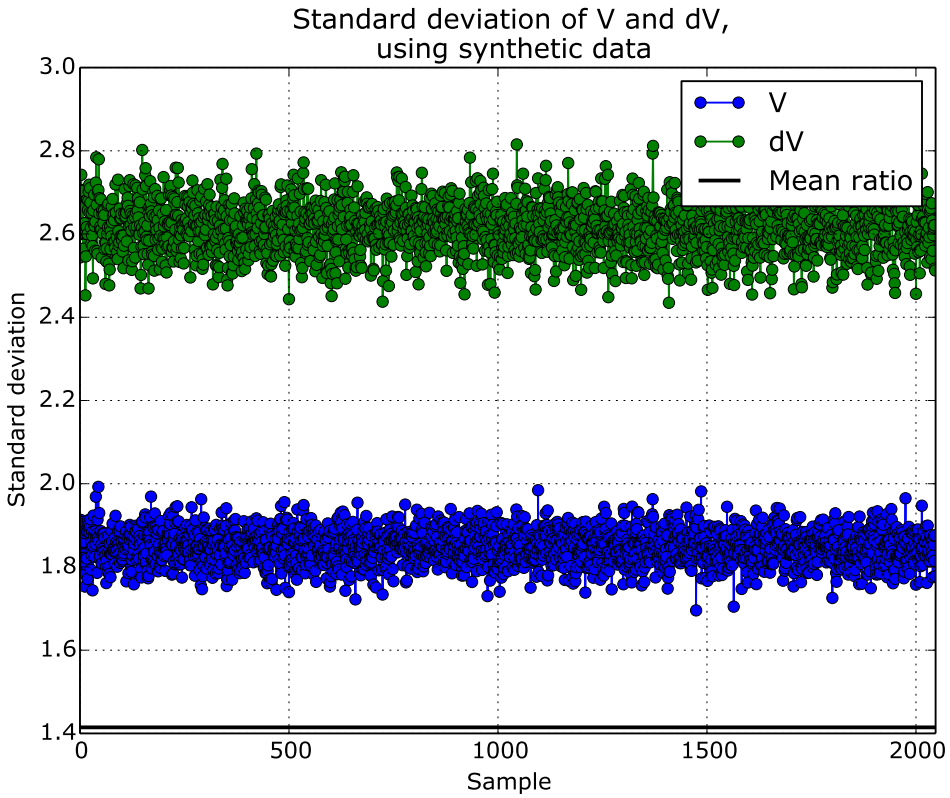

Figure A.2: Standard deviation of V and dV using synthetic data

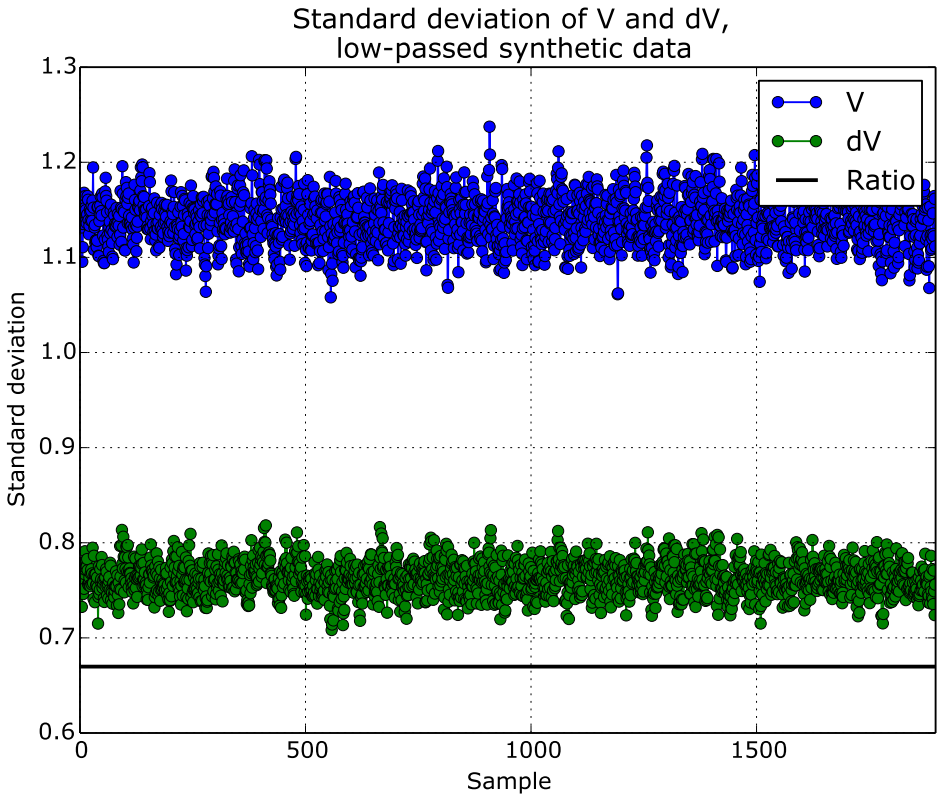

Figure A.3: The effect of lowpassing to the standard deviation, using synthetic data.

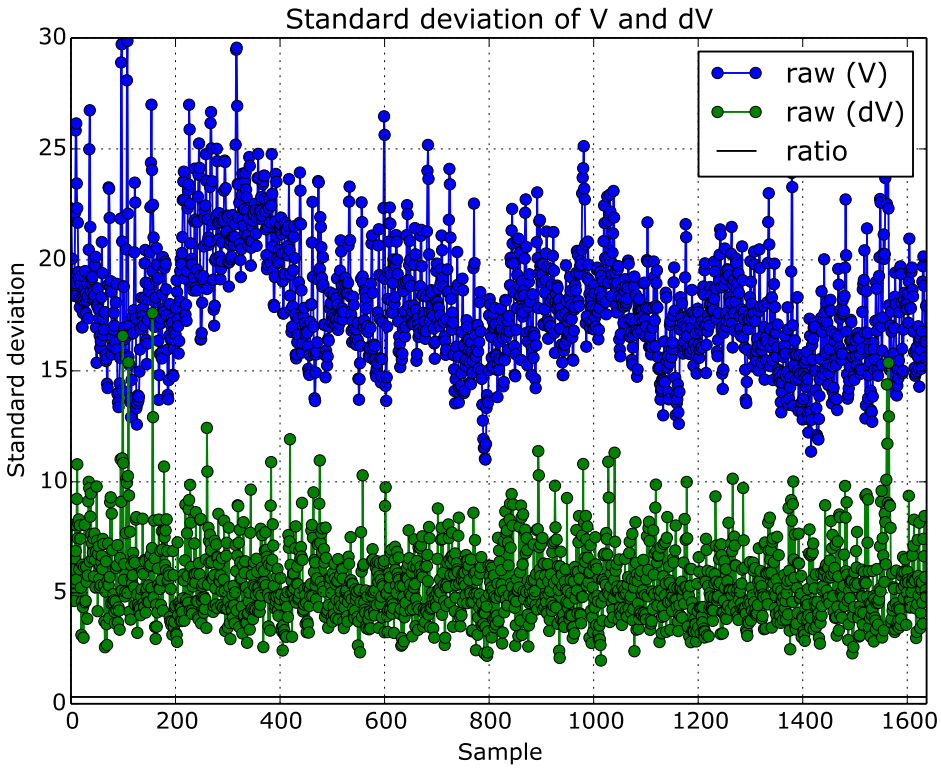

Figure A.4: Standard deviation calculation using raw data. Showing results for participant 1 at location CPz.

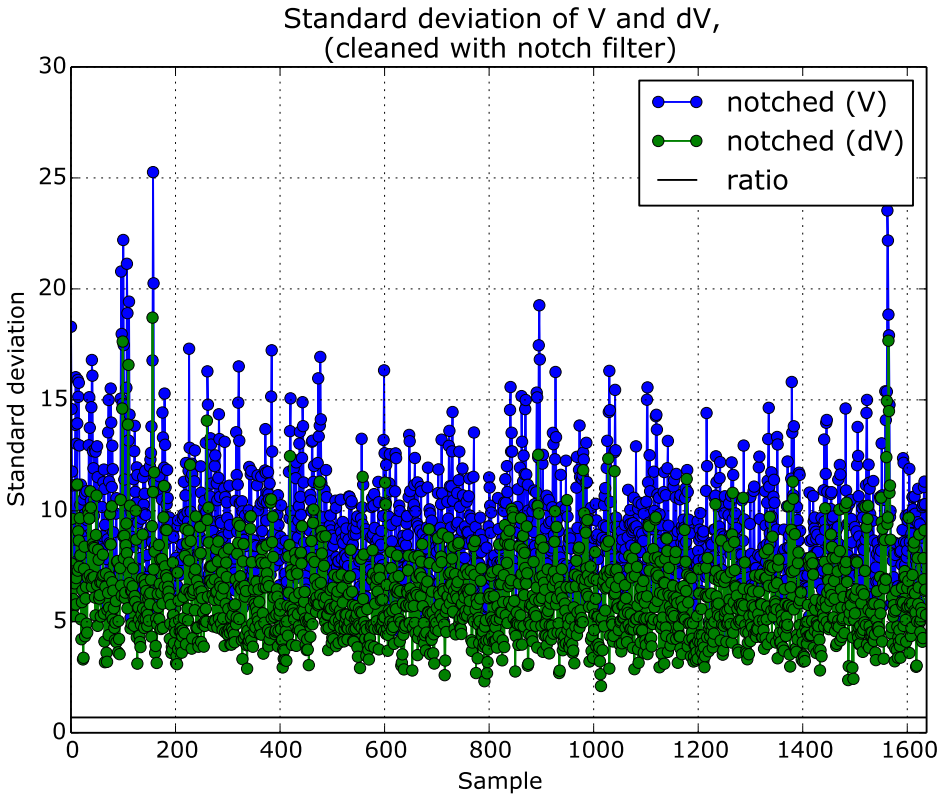

Figure A.5: Effects of the Notch filter on the standard deviation of dV and V. Data from participant 1 at location CPz.

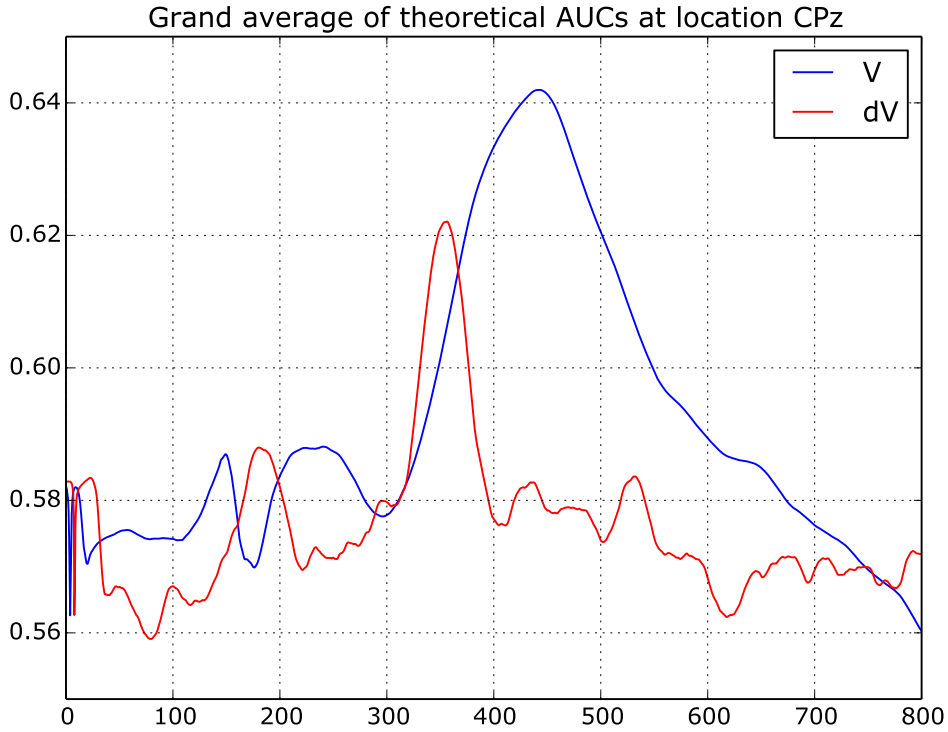

Figure A.6: Theoretical AUCs per sample of dV and V at electrode location CPz. Mean over trials across all participants is displayed.

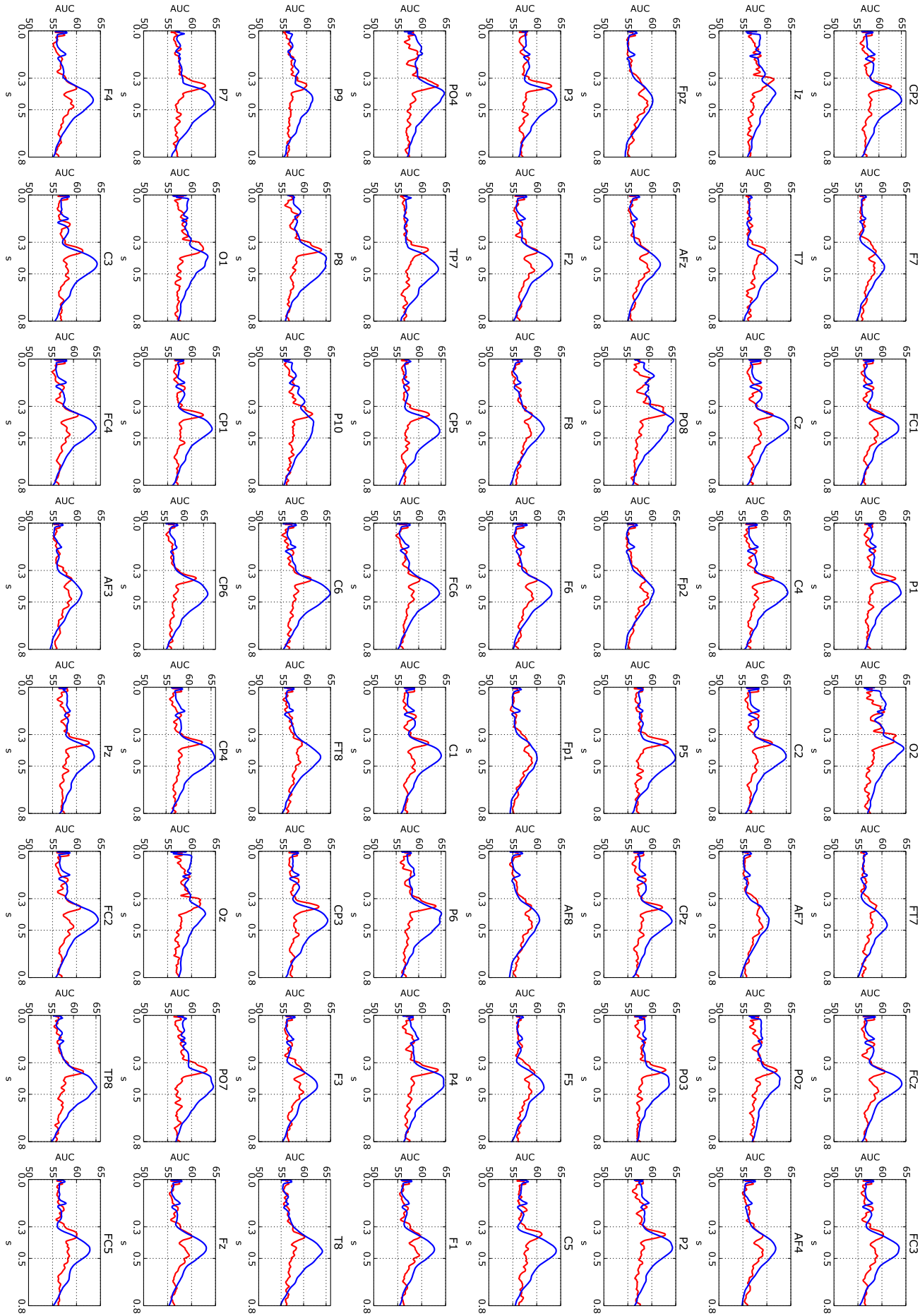

Figure A.7: Example of theoretical AUCs averaged across participants and direction runs, displaying results from all electrode sites.

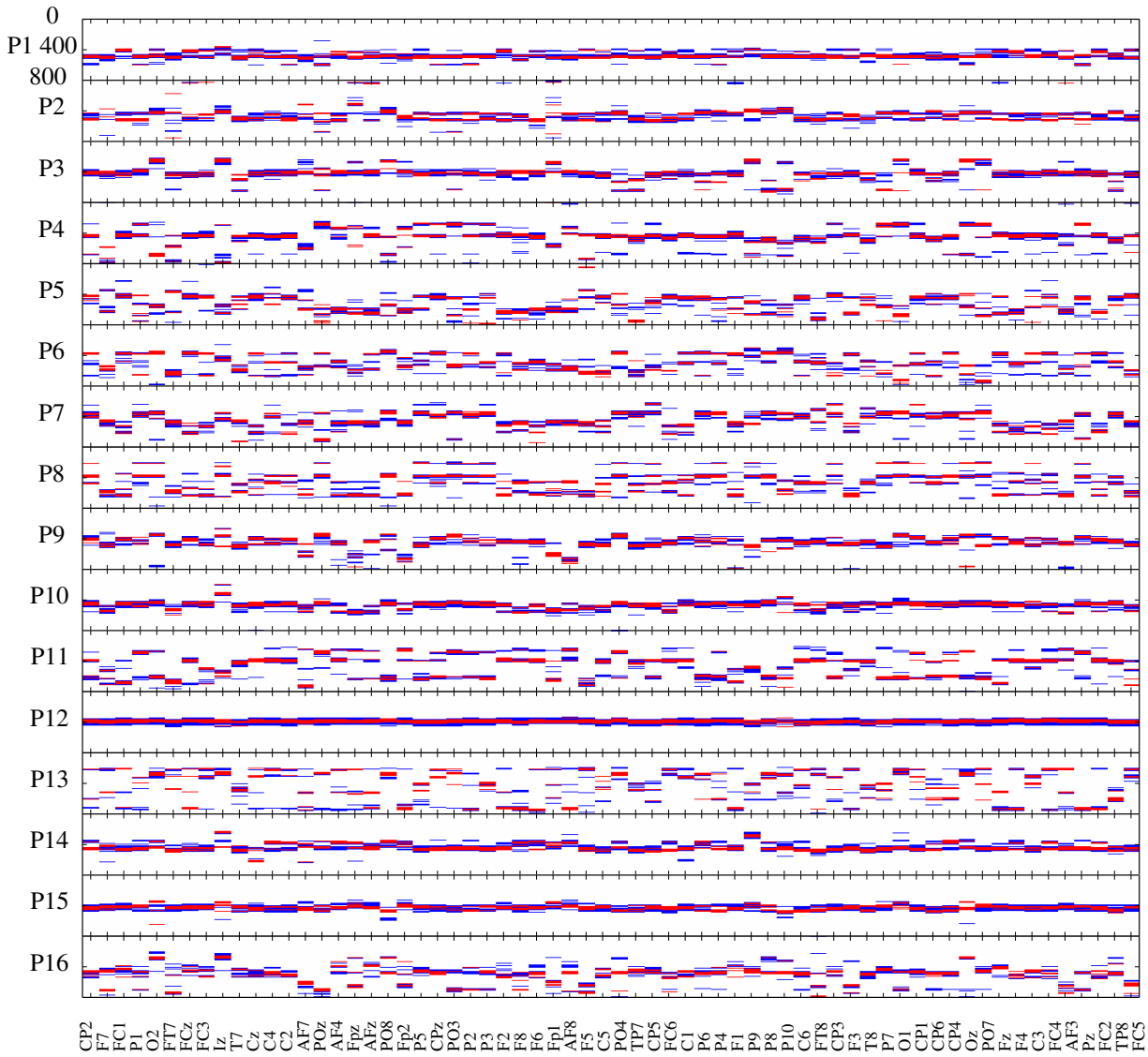

Figure A.8: Temporal differences of T and NT epochs according to ADM, V-based. Only the 24 selected samples per electrode are shown: the 12 best in red and the following 12 best in blue. The vertical position of each segment holds the time information (in ms) whereas the electrode locations are plotted on the horizontal axis.

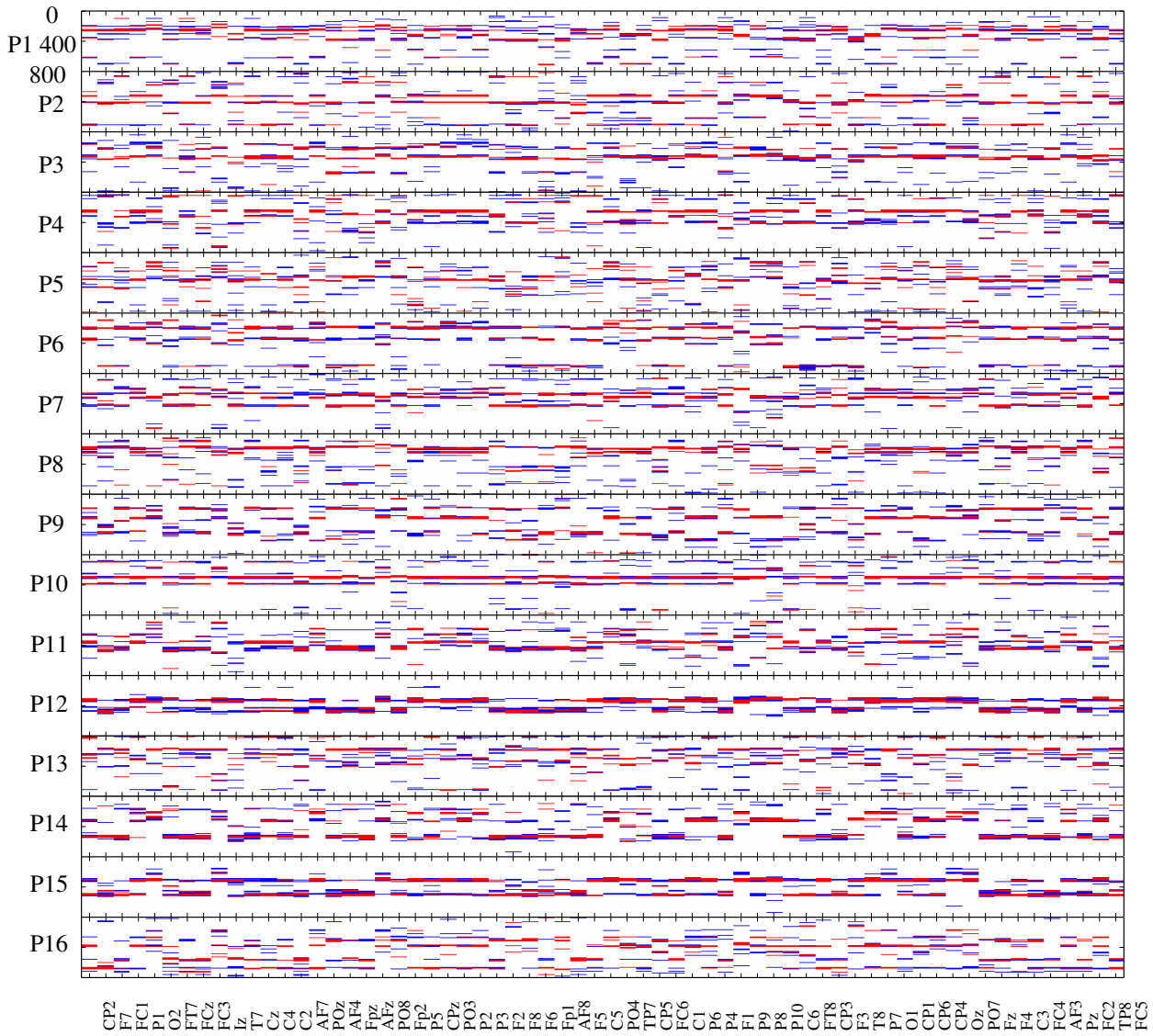

Figure A.9: Temporal differences of T and NT epochs according to ADM, dV-based. Only the 24 selected samples per electrode are shown: the 12 best in red and the following 12 best in blue. The vertical position of each segment holds the time information (in ms) whereas the electrode locations are plotted on the horizontal axis.

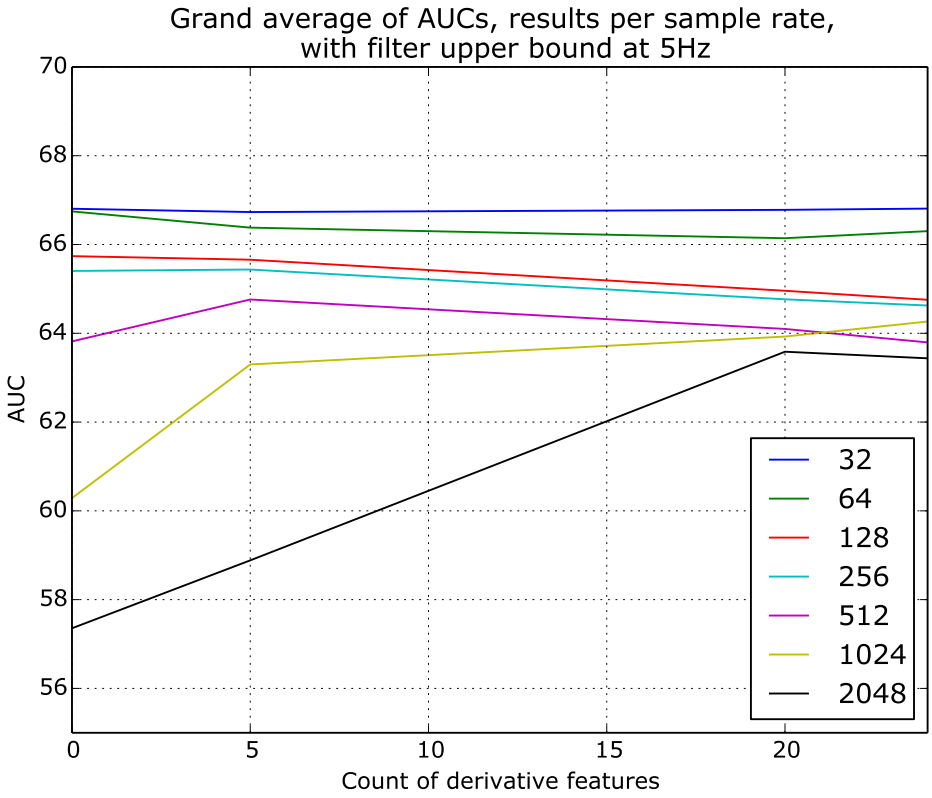

Figure A.10: Using $5Hz$ filter upper bound and a simple LDA for classification; averaged AUCs over participants for varying sampling rates

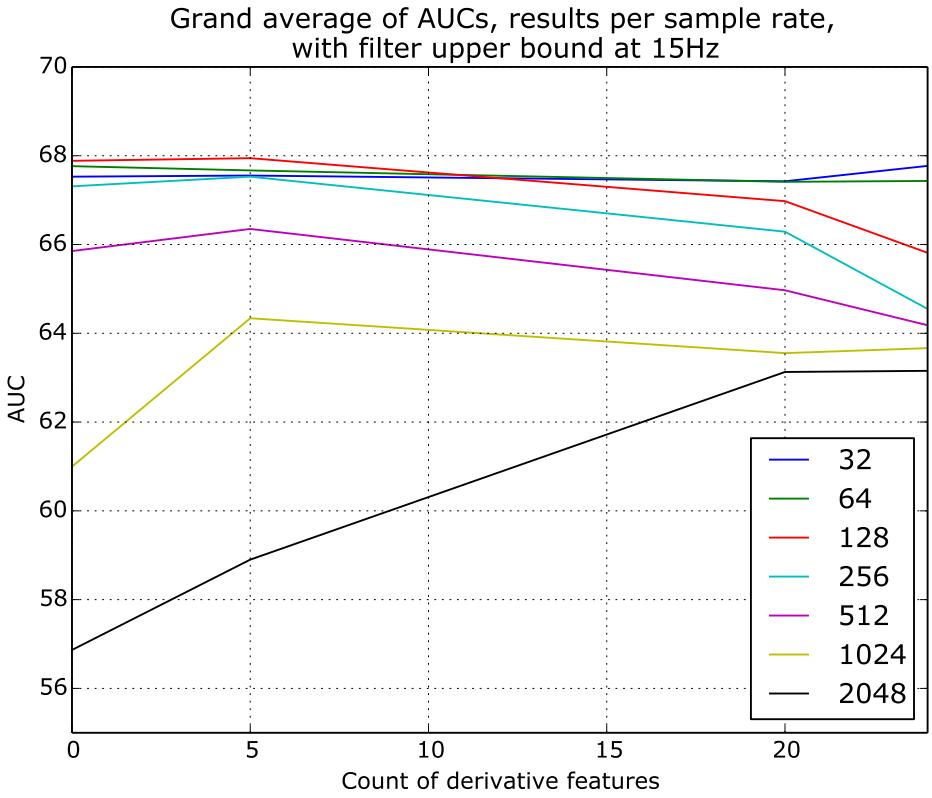

Figure A.11: Using $15Hz$ filter upper bound and a simple LDA for classification; averaged AUCs over participants for varying sampling rates

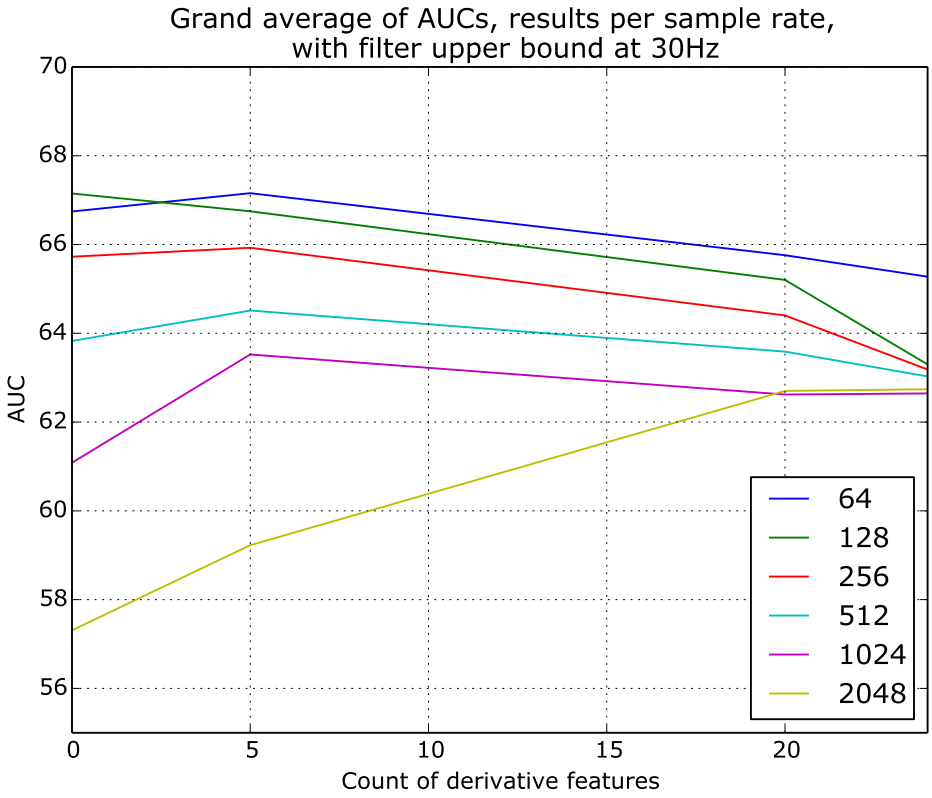

Figure A.12: Using 30Hz filter upper bound and a simple LDA for classification; averaged AUCs over participants for varying sampling rates

Appendix B

Clustering

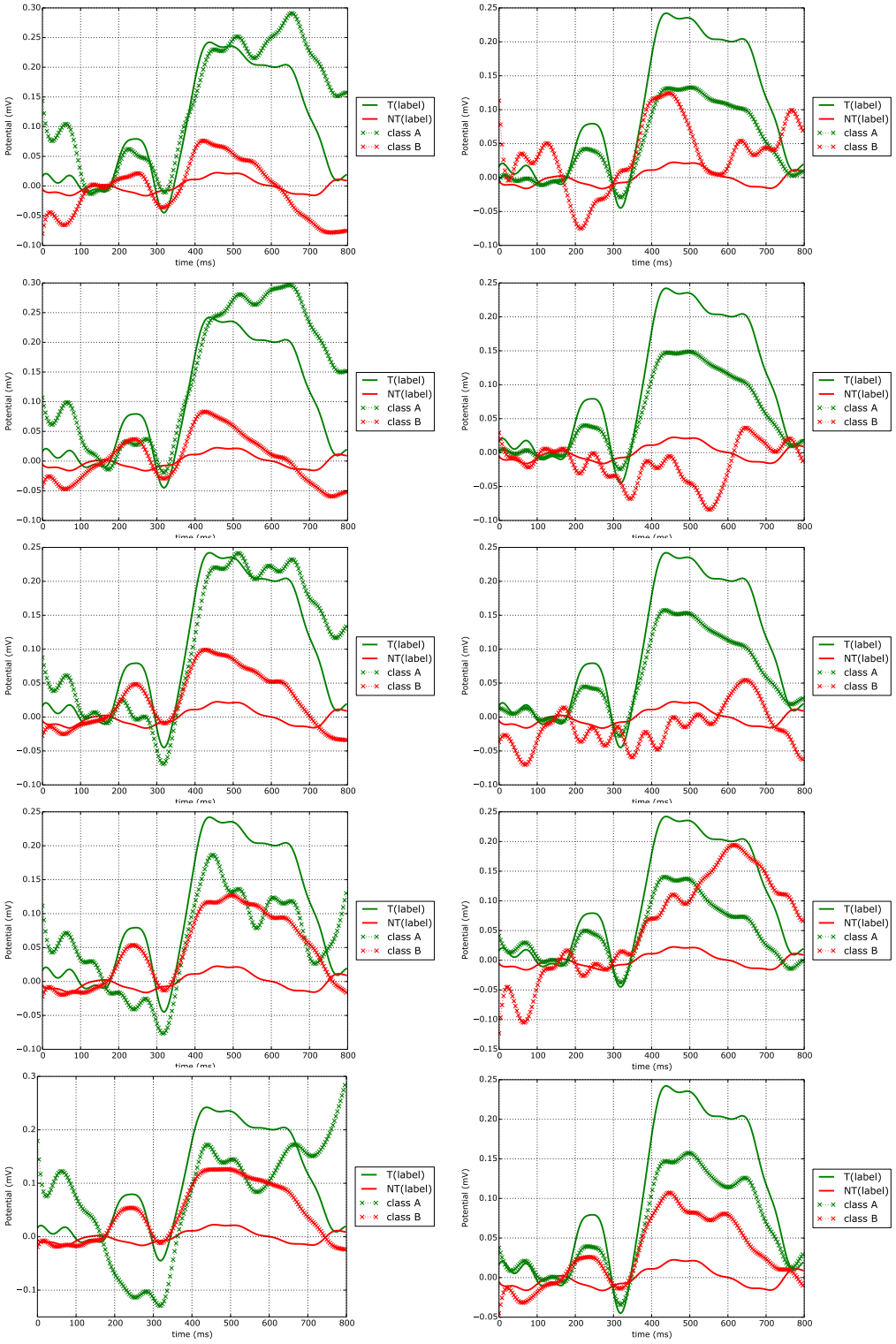

Figure B.1: Extracting the first group of epochs using the Kruskal-Walis test. On the left V, right dV. Showing results for participant 1

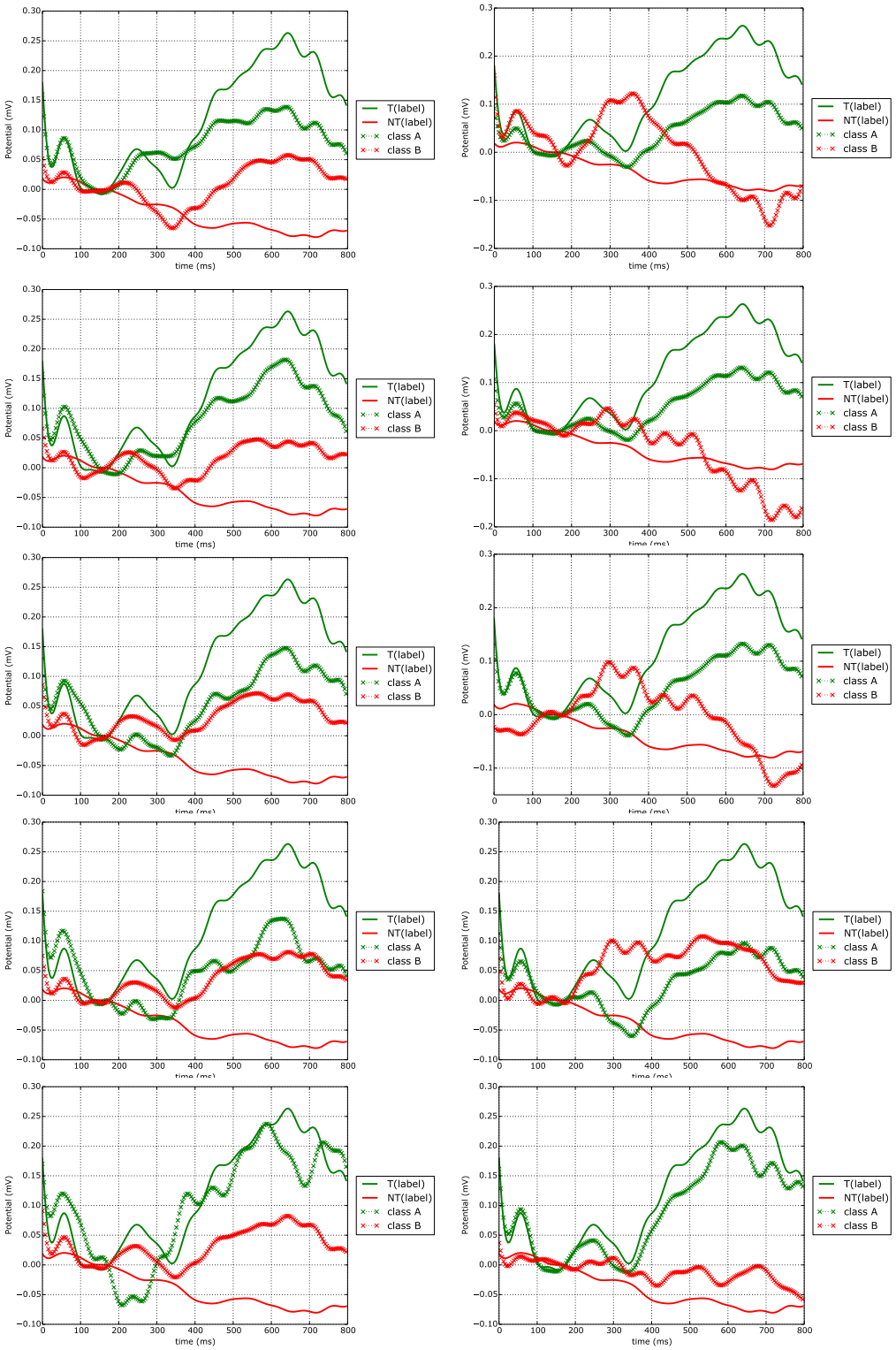

Figure B.2: Extracting the first group of epochs using the Kruskal-Walis test. On the left V, right dV. Showing results for participant 2

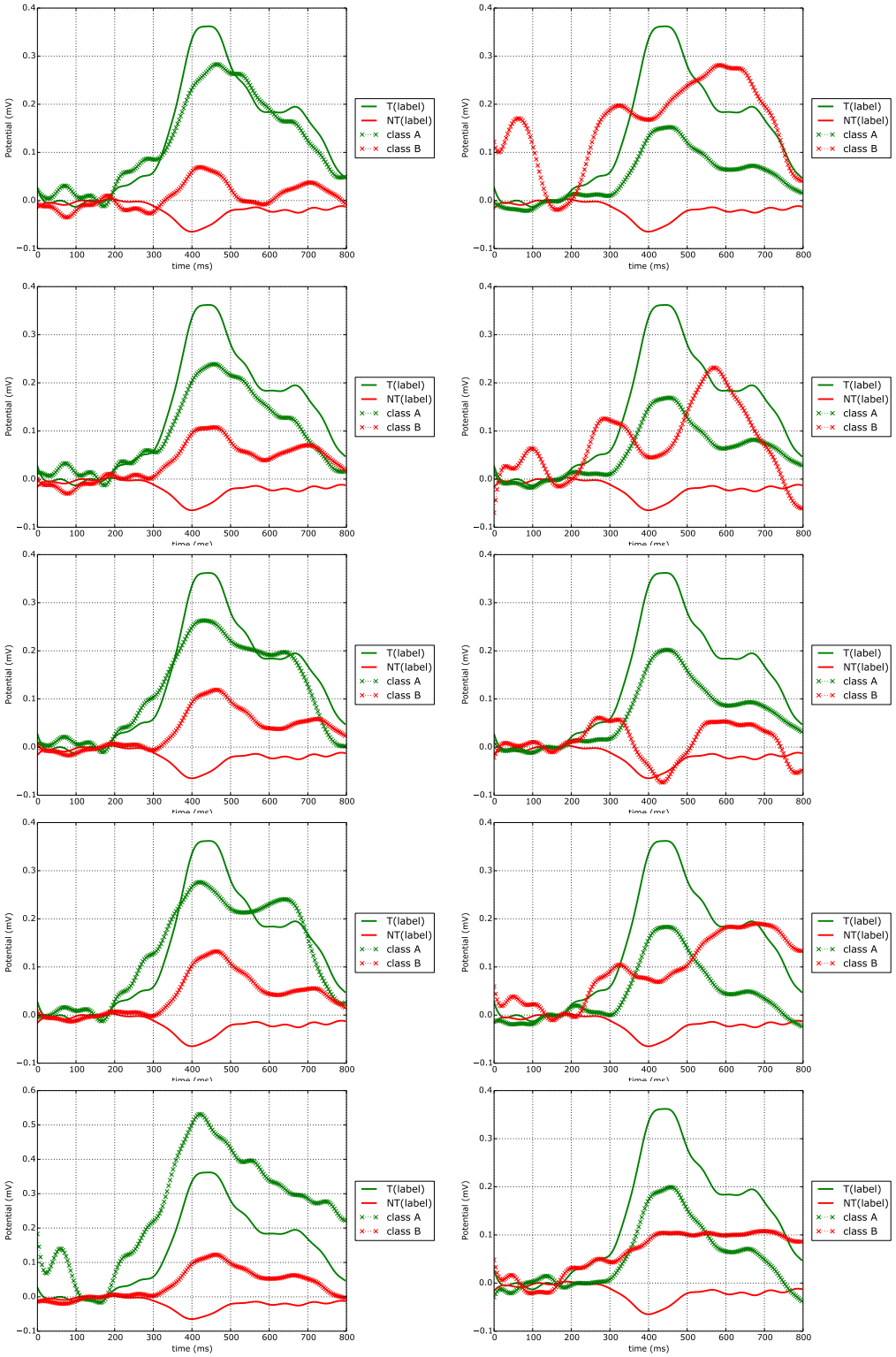

Figure B.3: Extracting the first group of epochs using the Kruskal-Walis test. On the left V, right dV. Showing results for participant 3

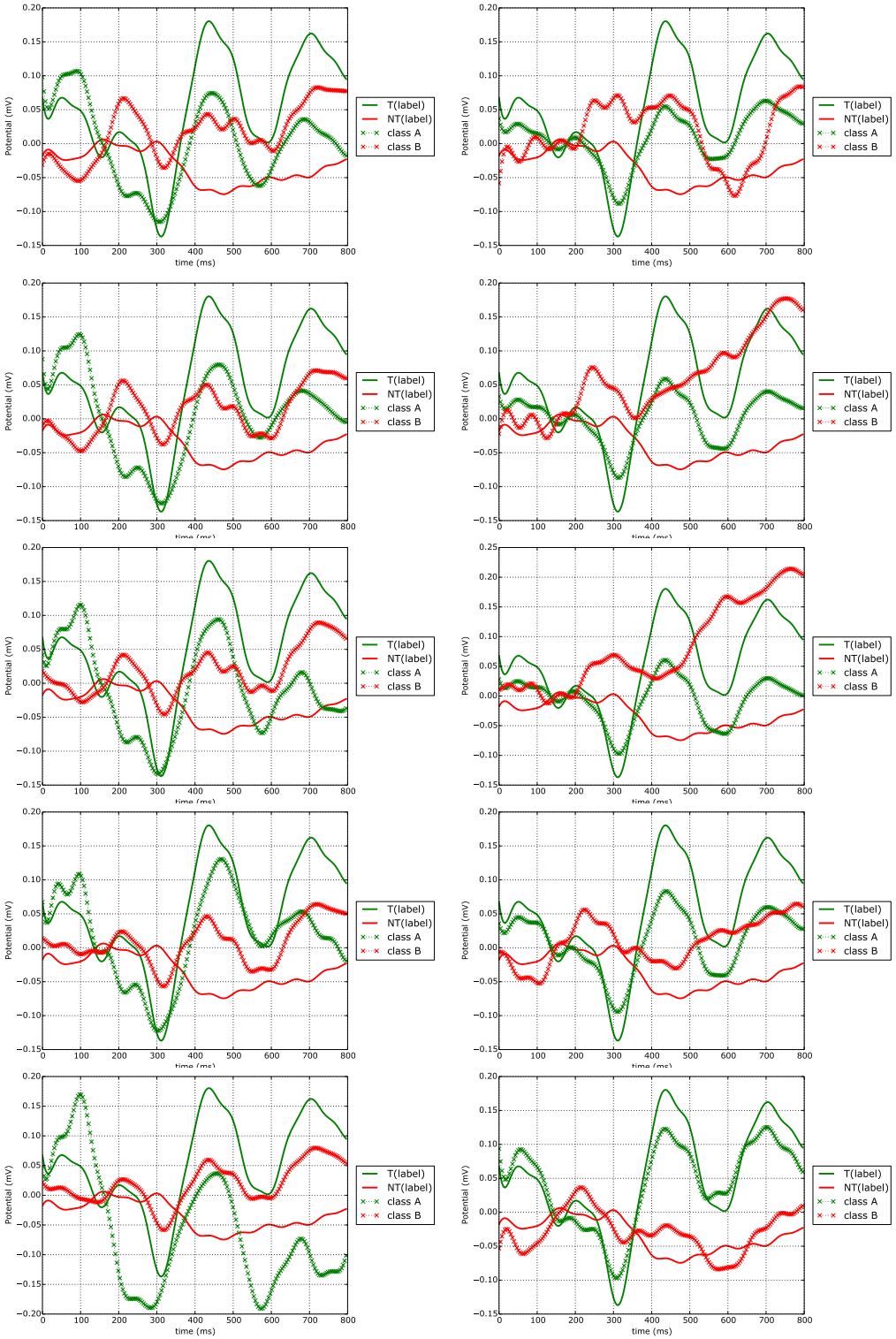

Figure B.4: Extracting the first group of epochs using the Kruskal-Walis test. On the left V, right dV. Showing results for participant 4

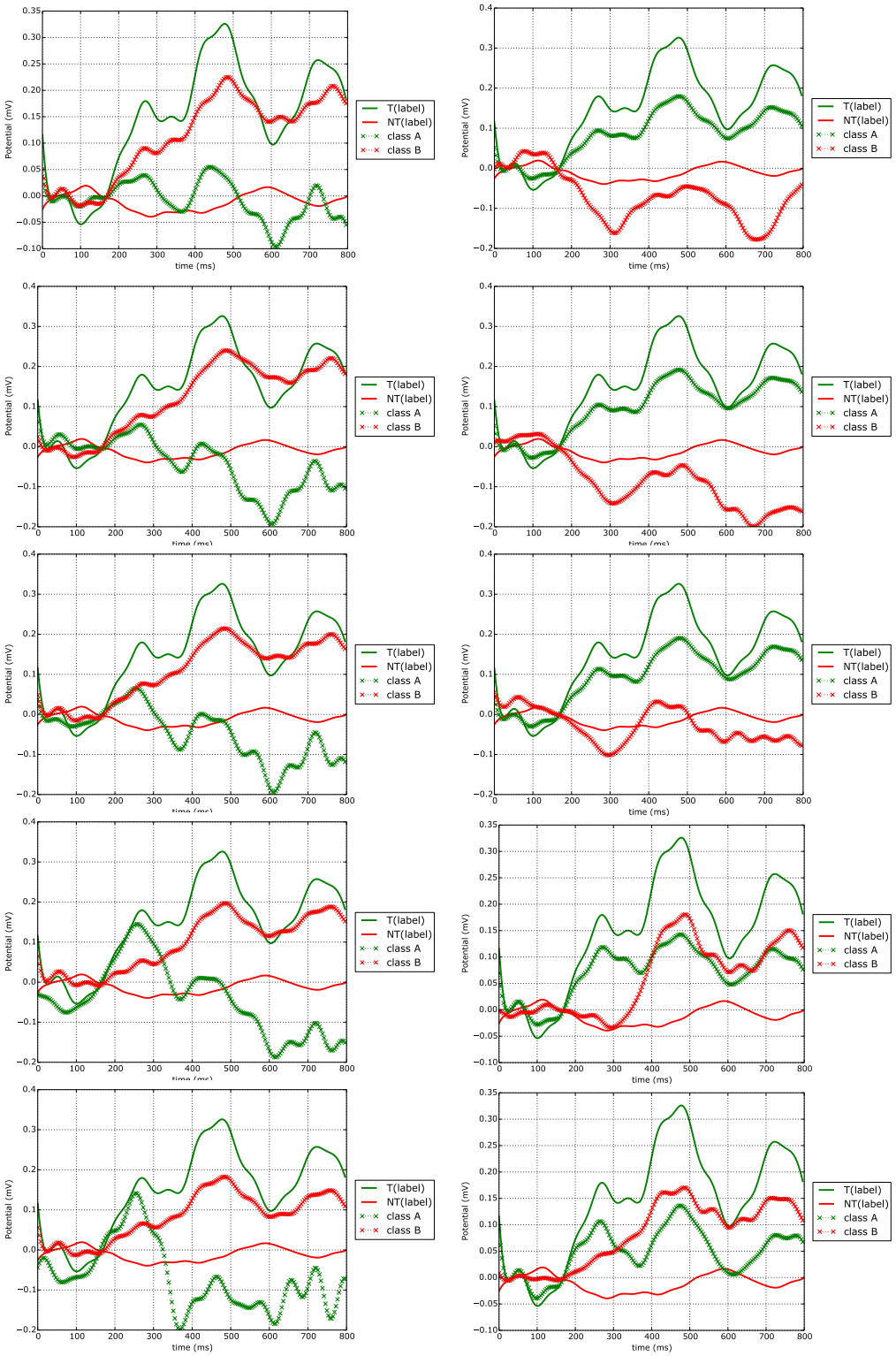

Figure B.5: Extracting the first group of epochs using the Kruskal-Walis test. On the left V, right dV. Showing results for participant 5

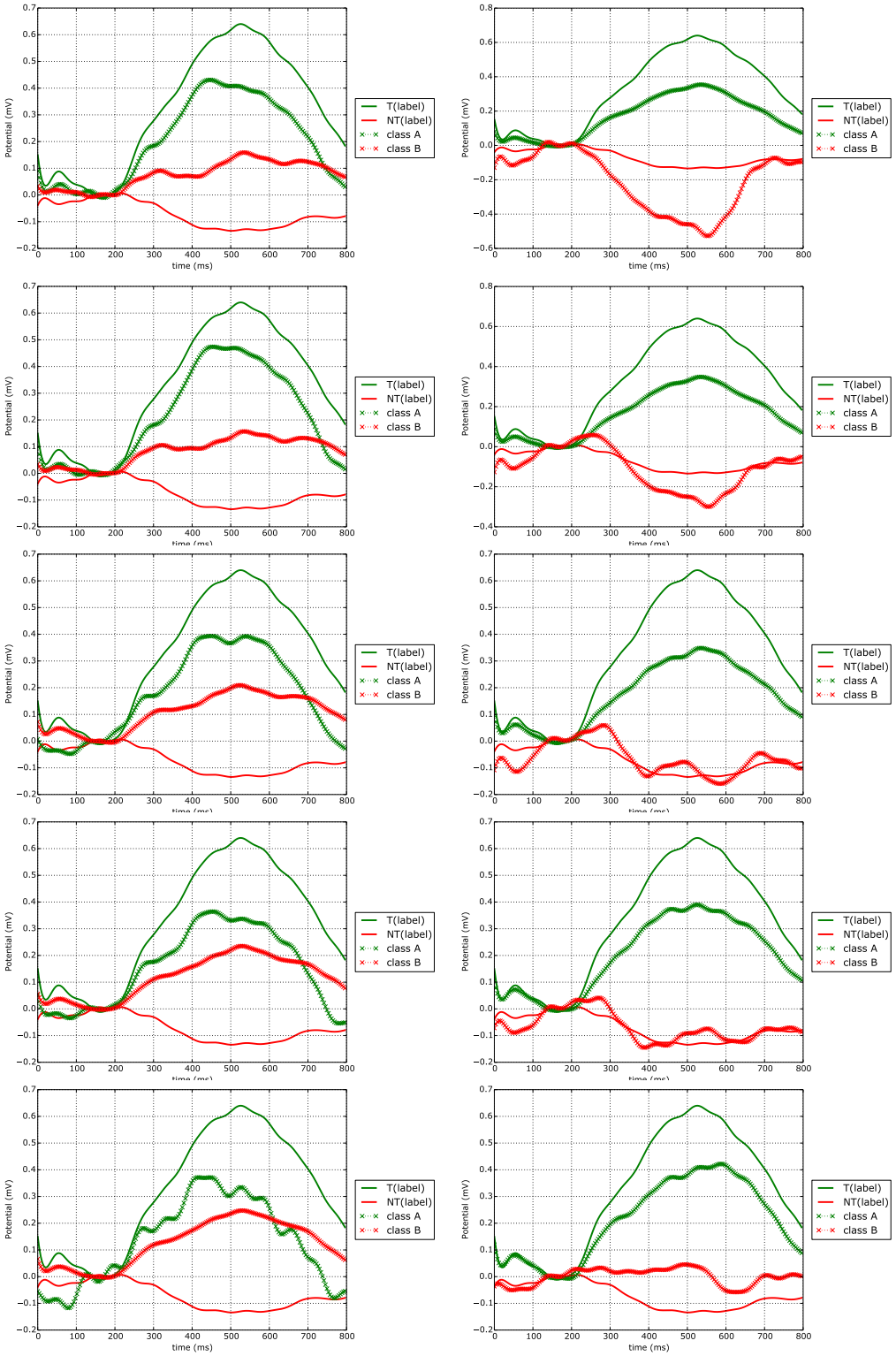

Figure B.6: Extracting the first group of epochs using the Kruskal-Walis test. On the left V, right dV. Showing results for participant 6

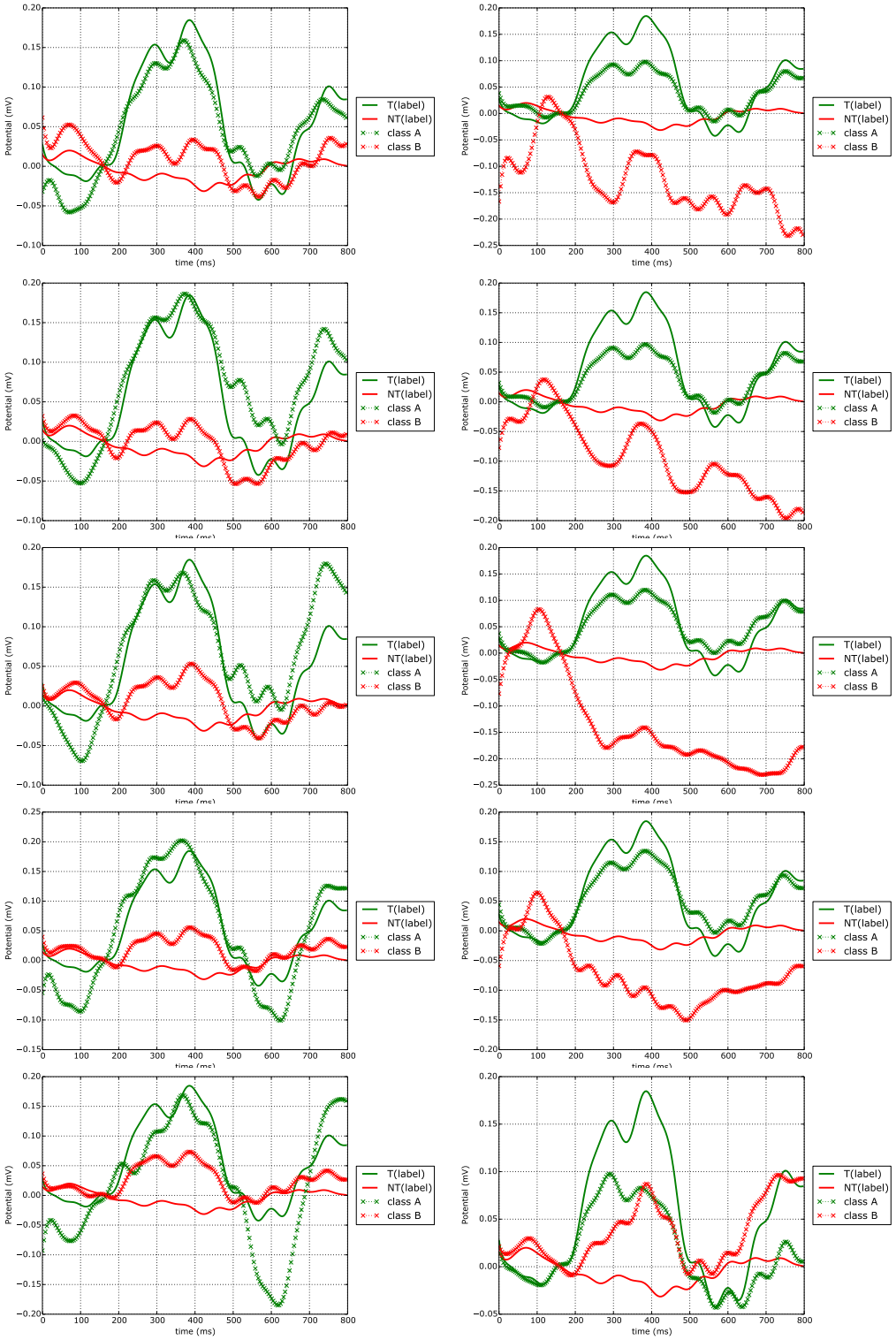

Figure B.7: Extracting the first group of epochs using the Kruskal-Walis test. On the left V, right dV. Showing results for participant 7

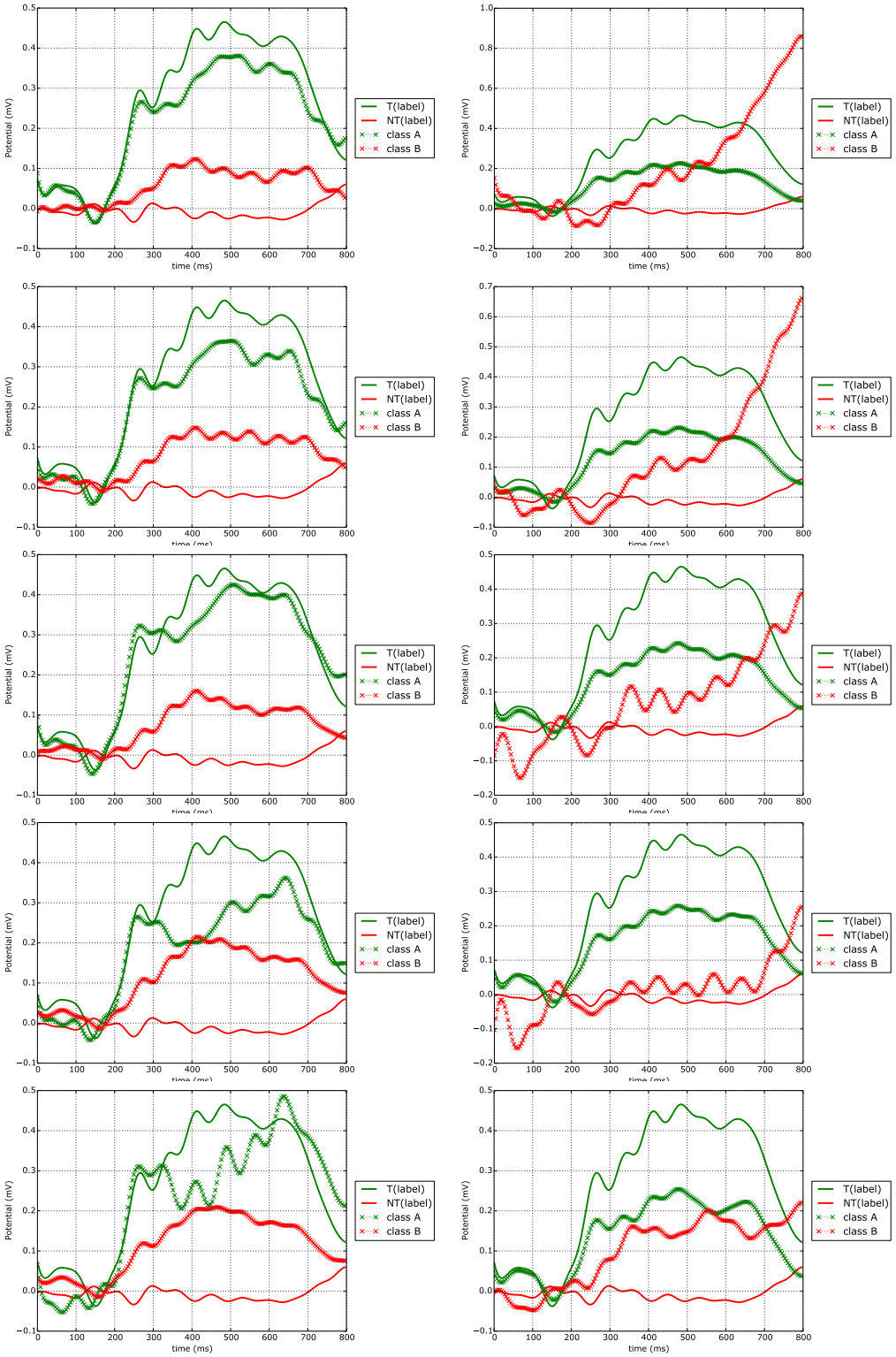

Figure B.8: Extracting the first group of epochs using the Kruskal-Walis test. On the left V, right dV. Showing results for participant 8

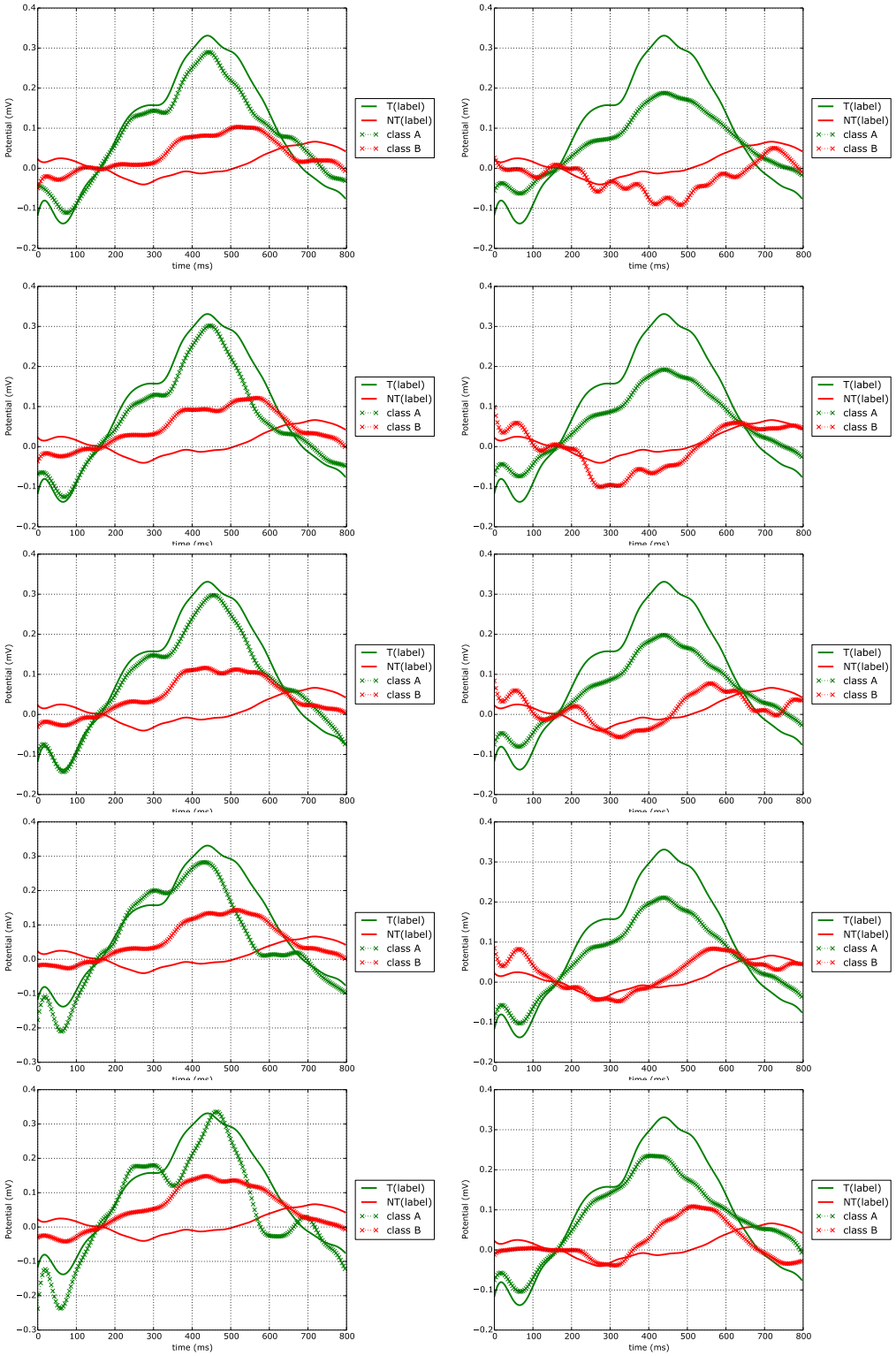

Figure B.9: Extracting the first group of epochs using the Kruskal-Walis test. On the left V, right dV. Showing results for participant 9

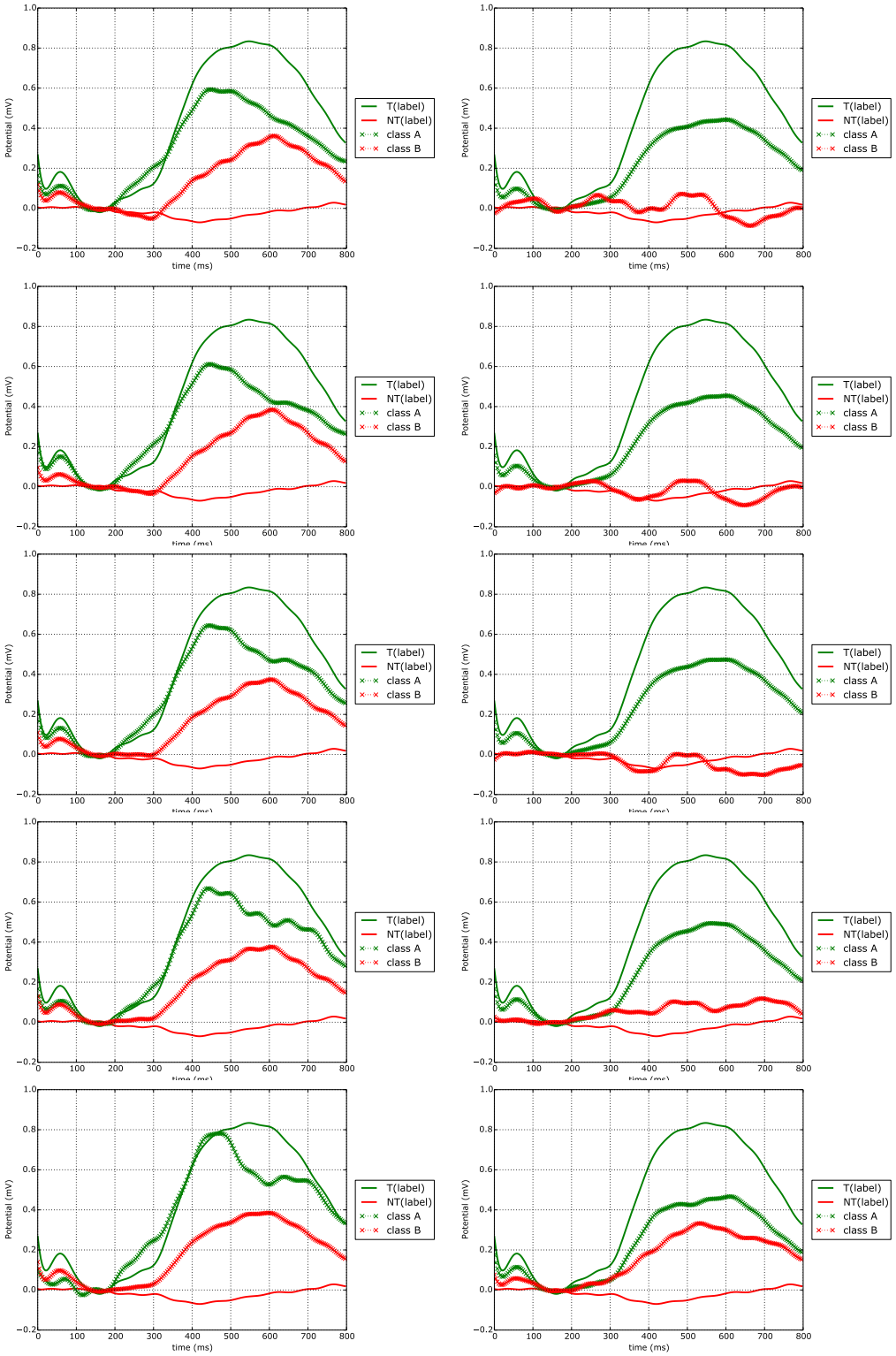

Figure B.10: Extracting the first group of epochs using the Kruskal-Walis test. On the left V, right dV. Showing results for participant 10

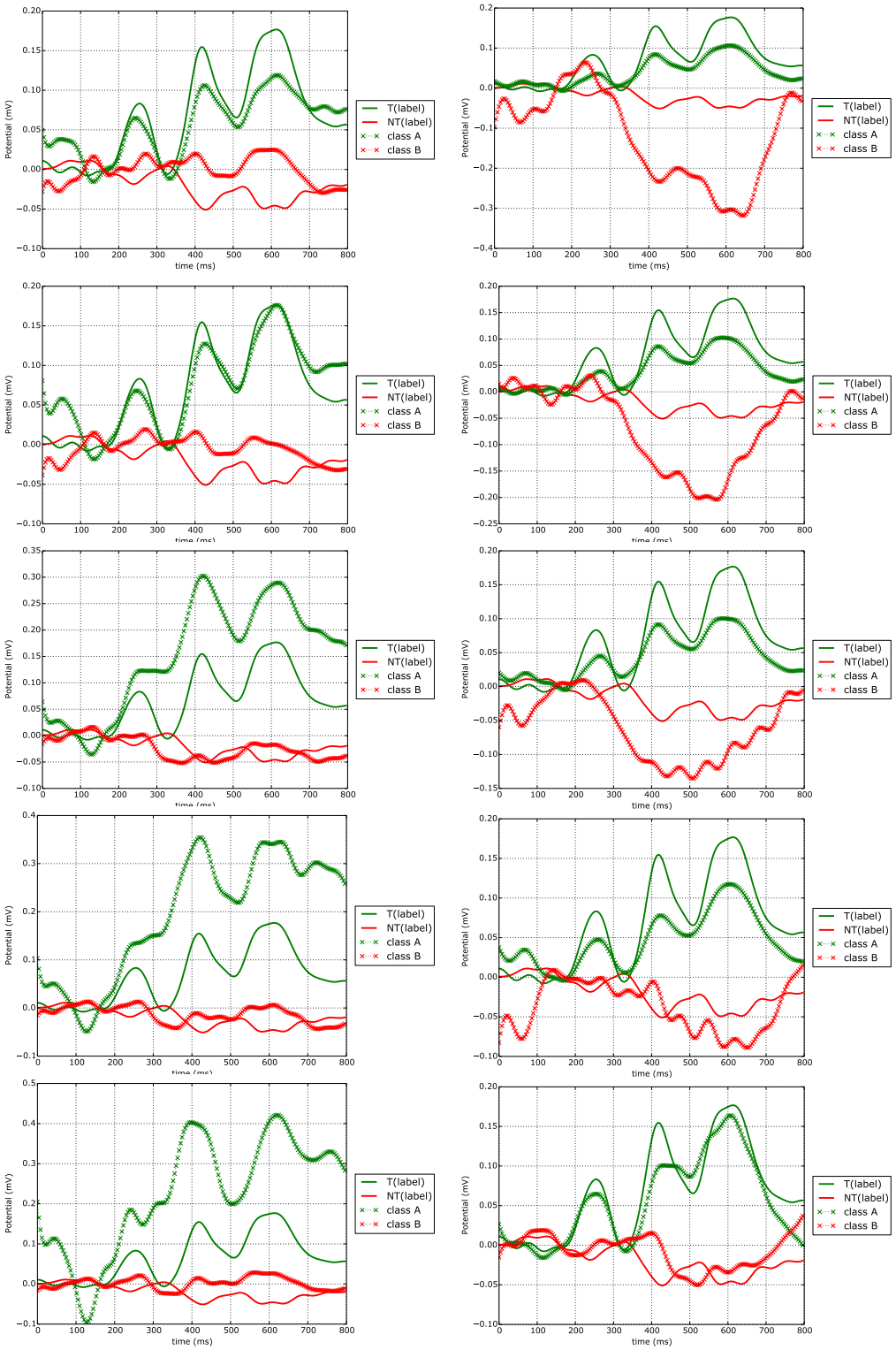

Figure B.11: Extracting the first group of epochs using the Kruskal-Walis test. On the left V, right dV. Showing results for participant 11

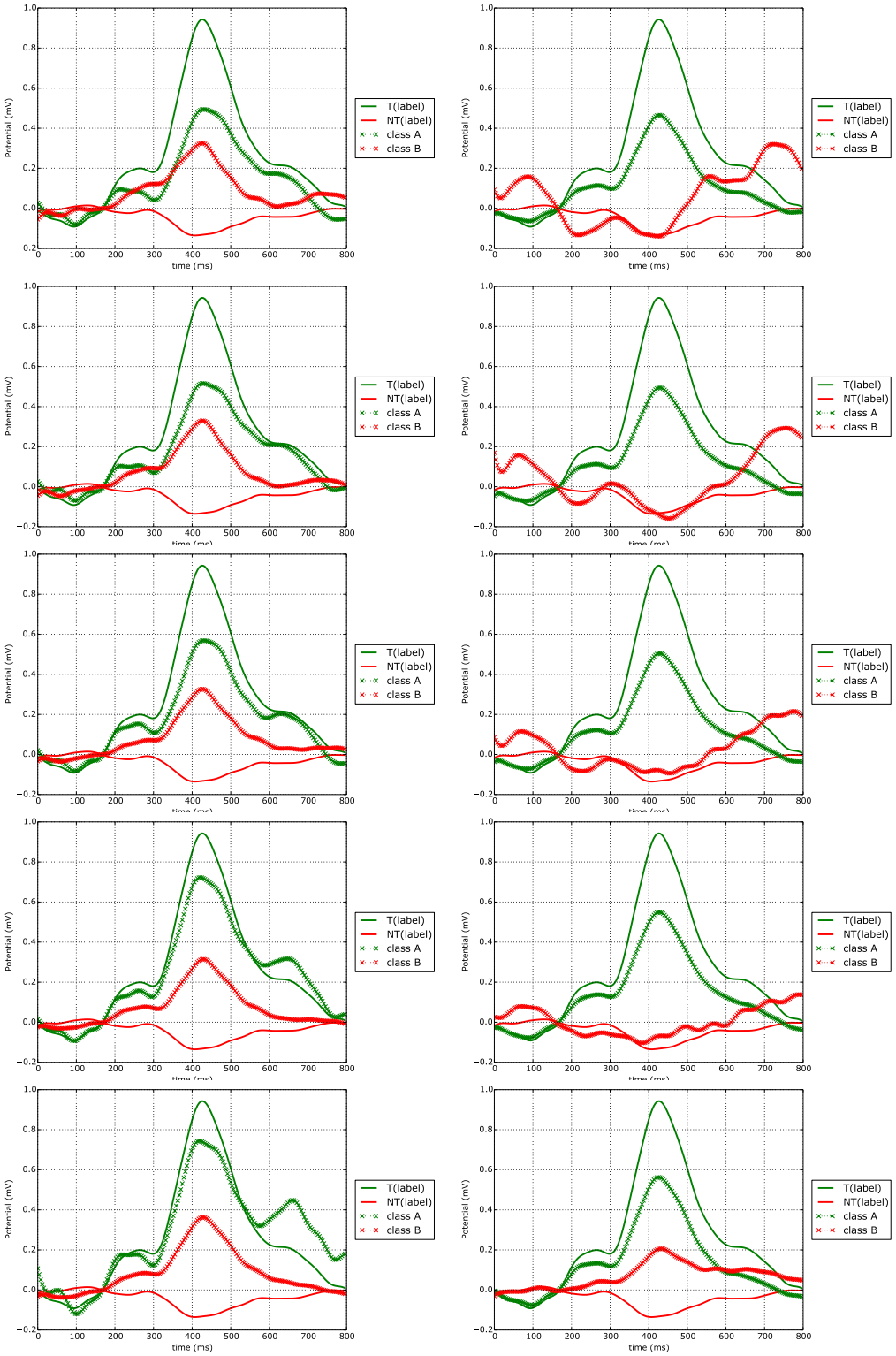

Figure B.12: Extracting the first group of epochs using the Kruskal-Walis test. On the left V, right dV. Showing results for participant 12

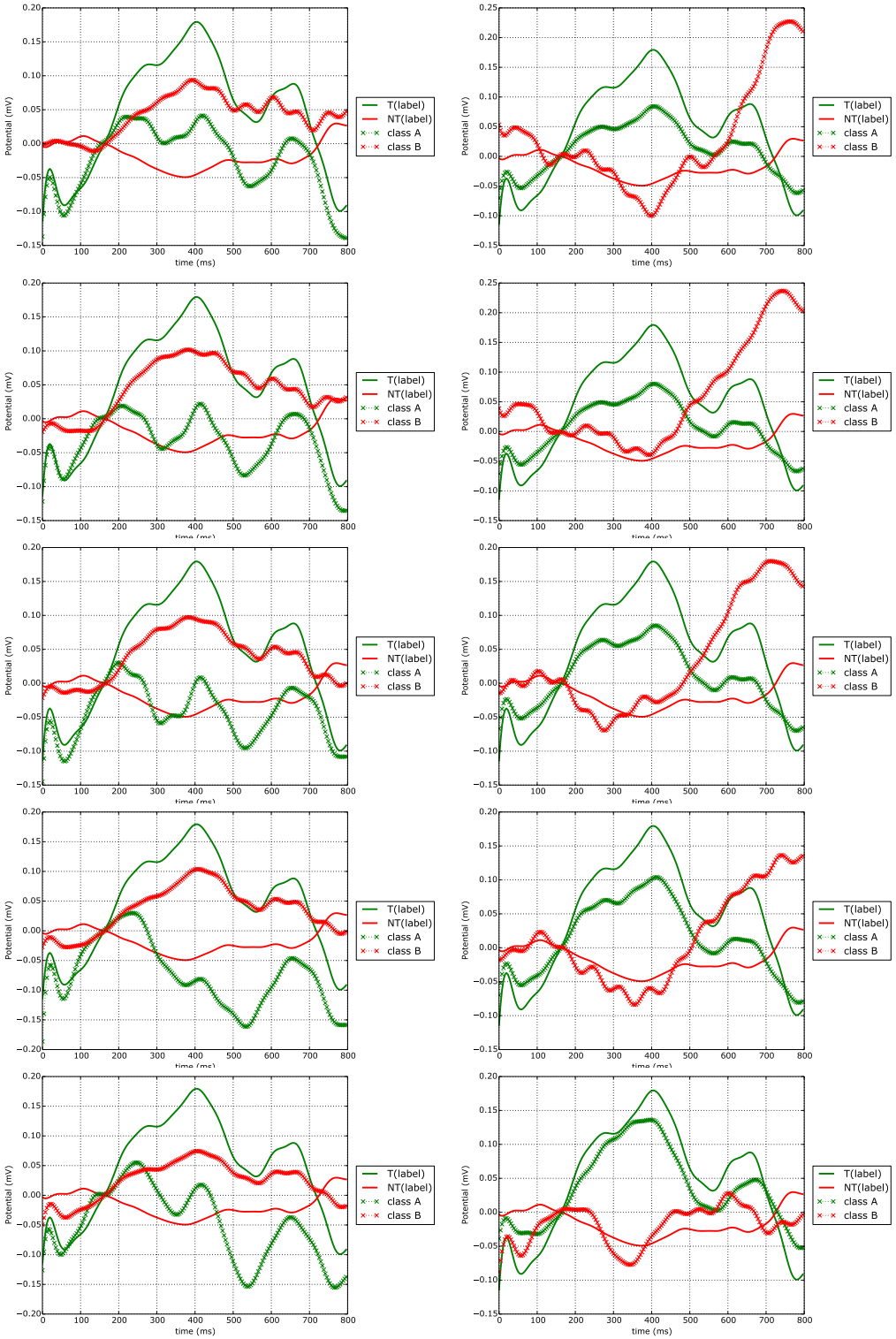

Figure B.13: Extracting the first group of epochs using the Kruskal-Walis test. On the left V, right dV. Showing results for participant 13

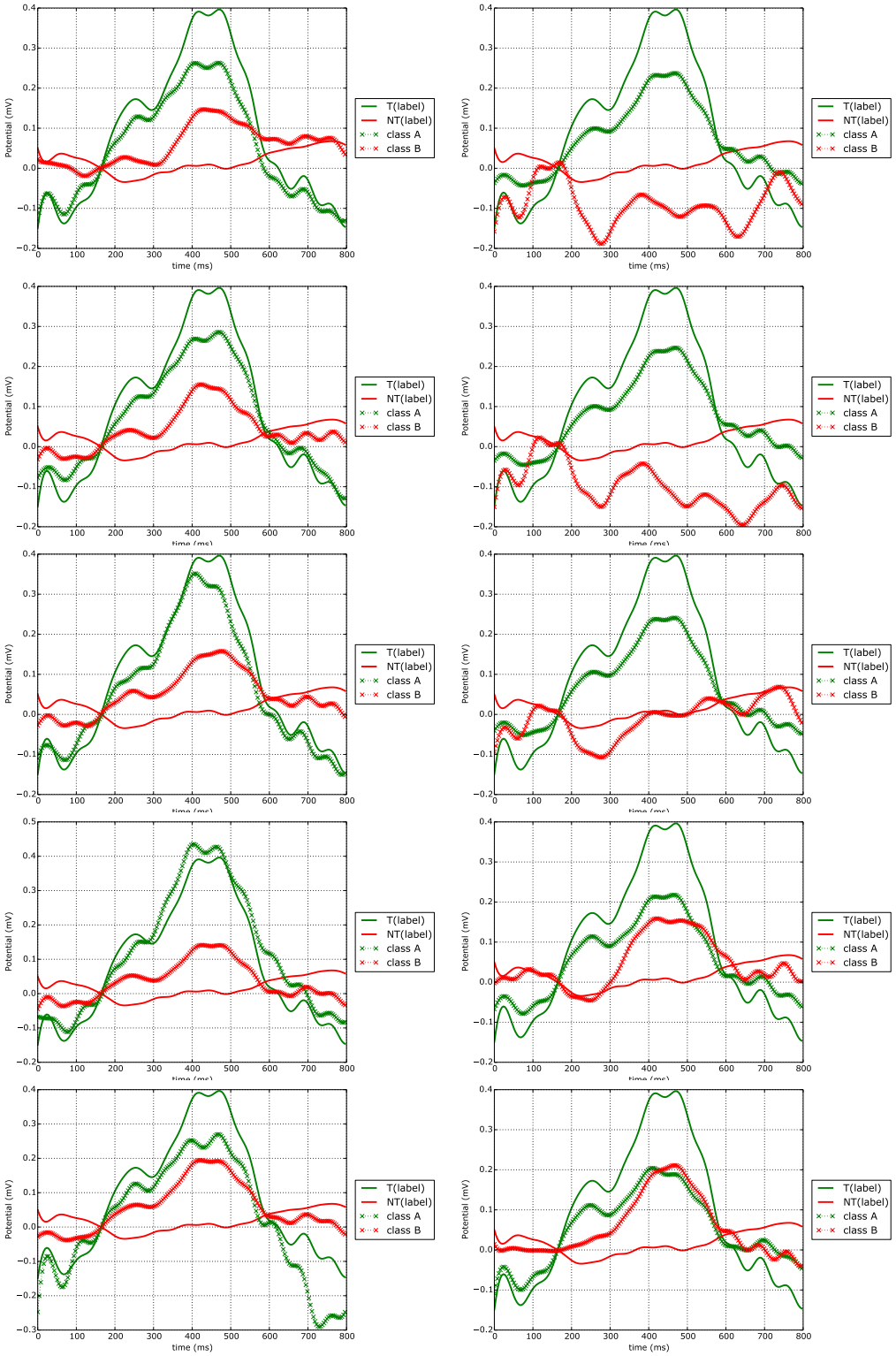

Figure B.14: Extracting the first group of epochs using the Kruskal-Walis test. On the left V, right dV. Showing results for participant 14

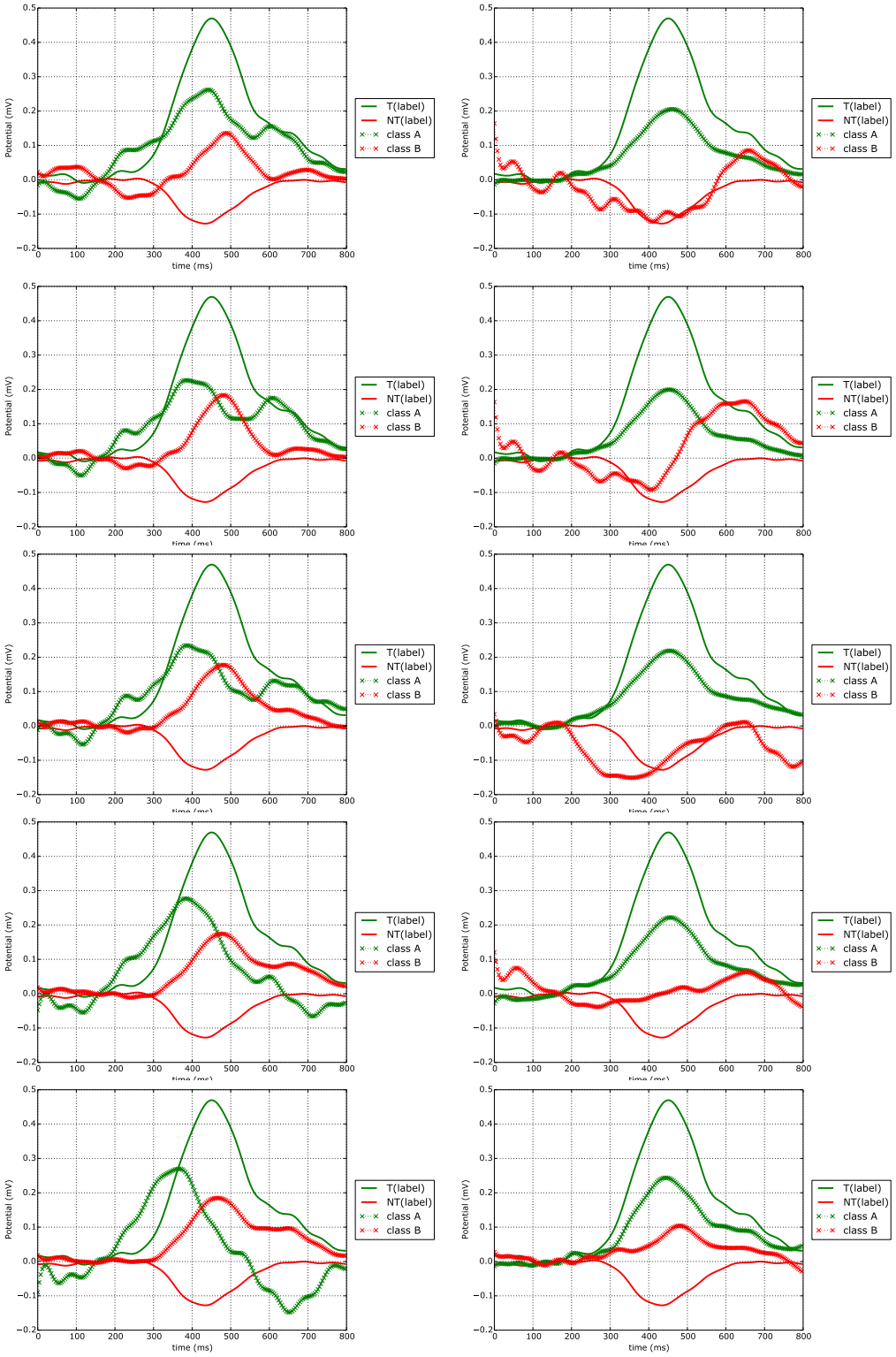

Figure B.15: Extracting the first group of epochs using the Kruskal-Walis test. On the left V, right dV. Showing results for participant 15

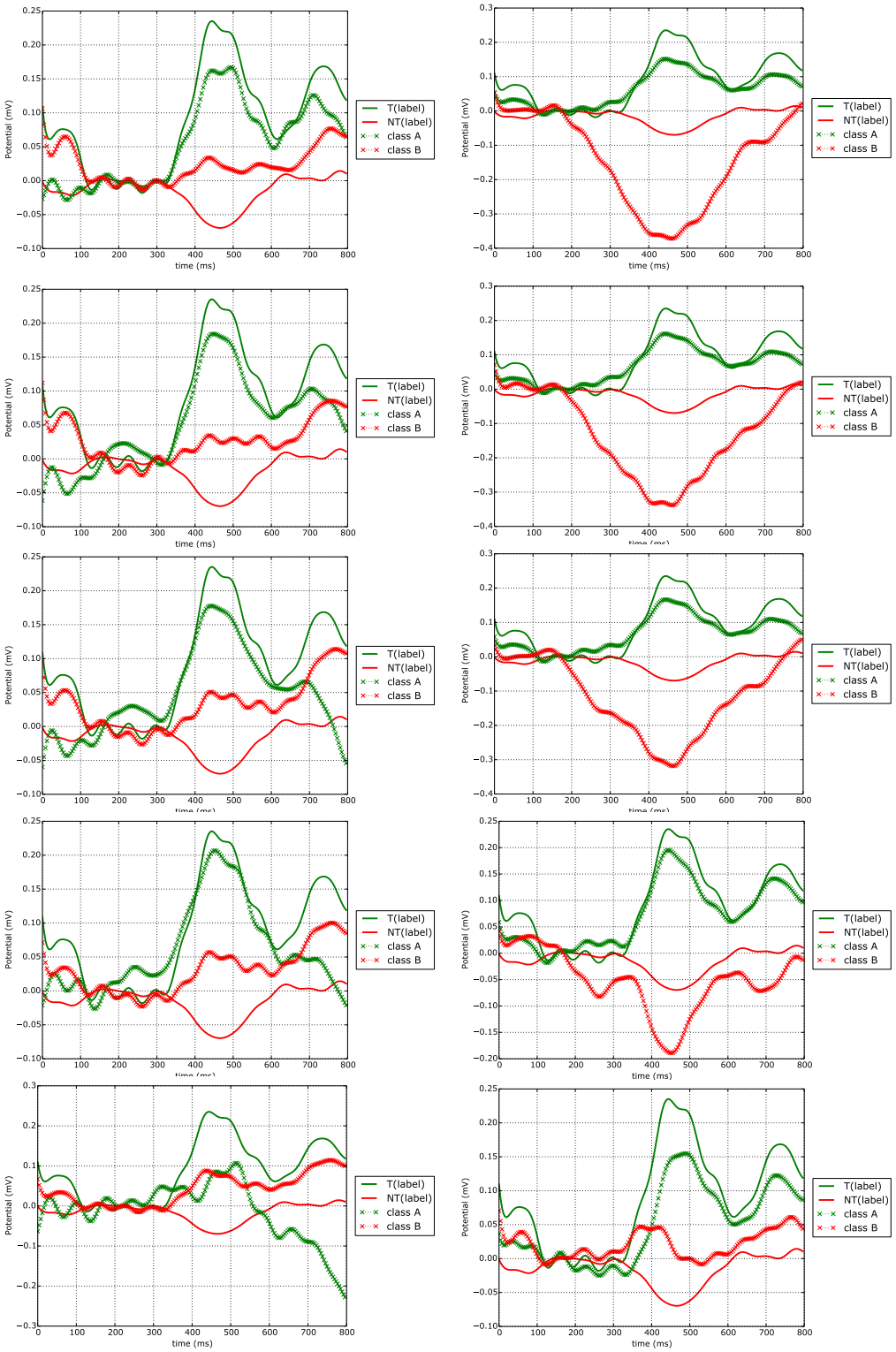

Figure B.16: Extracting the first group of epochs using the Kruskal-Walis test. On the left V, right dV. Showing results for participant 16

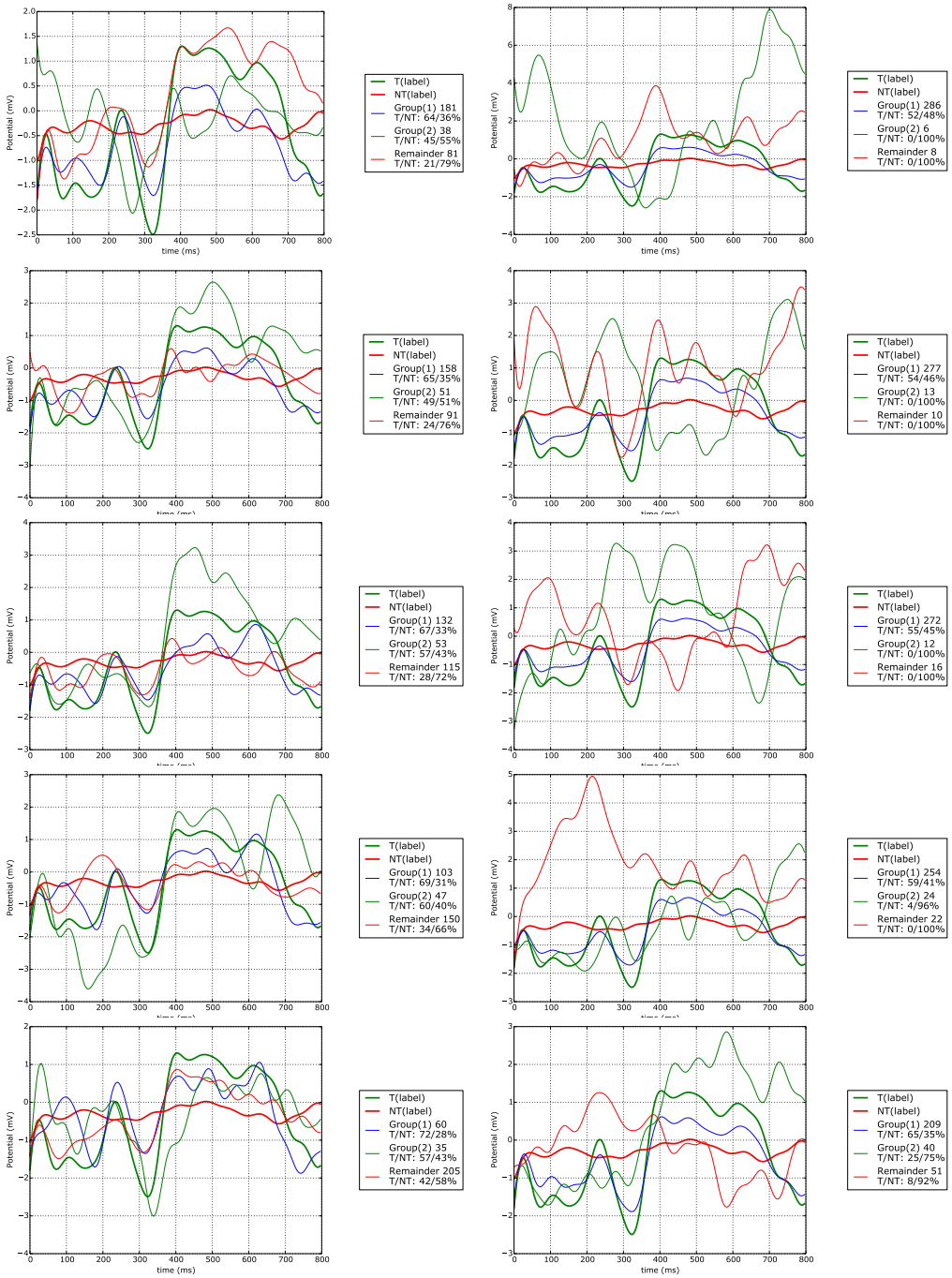

Figure B.17: Extracting two groups of epochs using the Kruskal-Walis test. On the left V, right dV. Showing results for participant 1

Figure B.18: Extracting two groups of epochs using the Kruskal-Walis test. On the left V, right dV. Showing results for participant 2

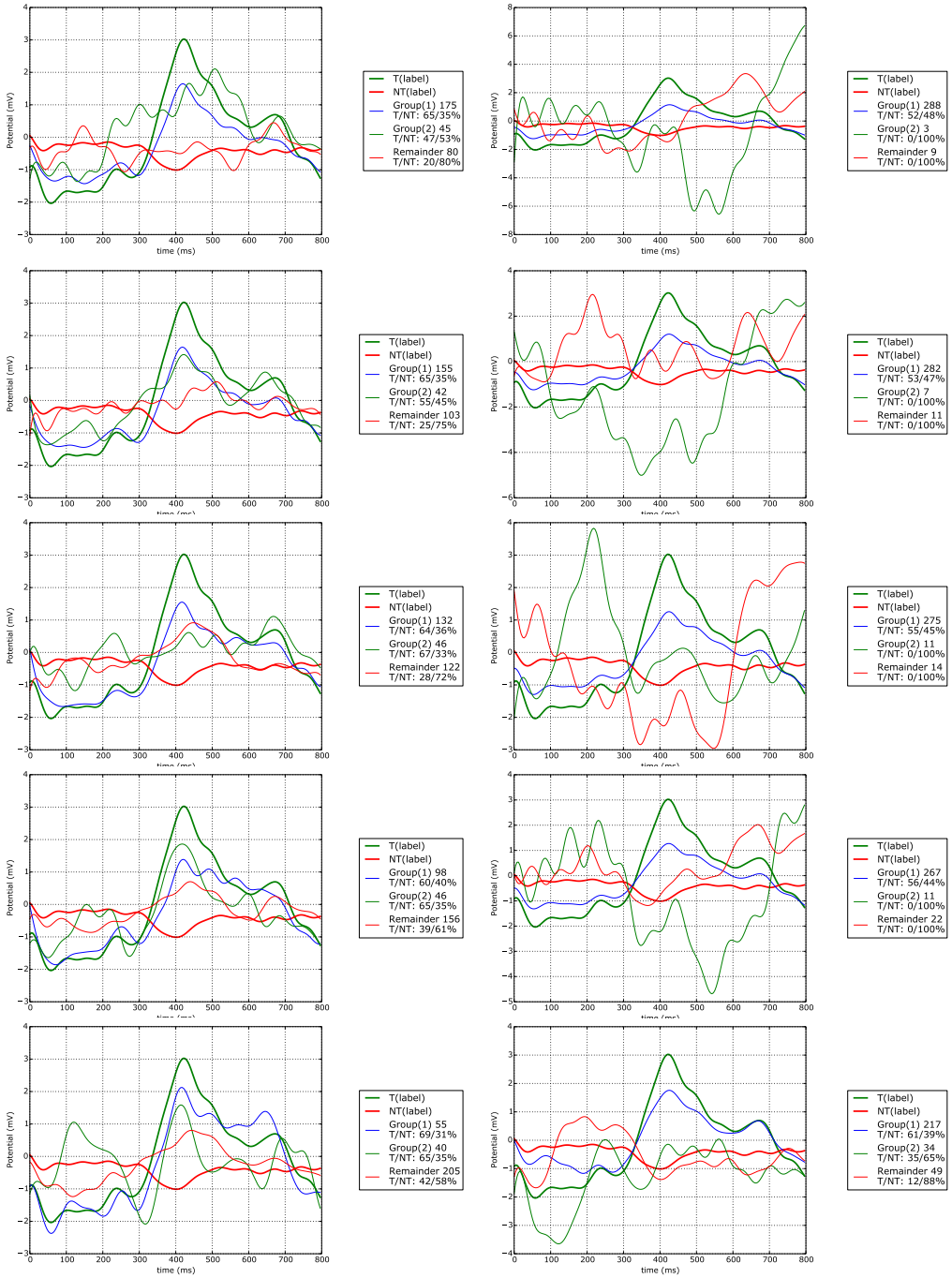

Figure B.19: Extracting two groups of epochs using the Kruskal-Walis test. On the left V, right dV. Showing results for participant 3

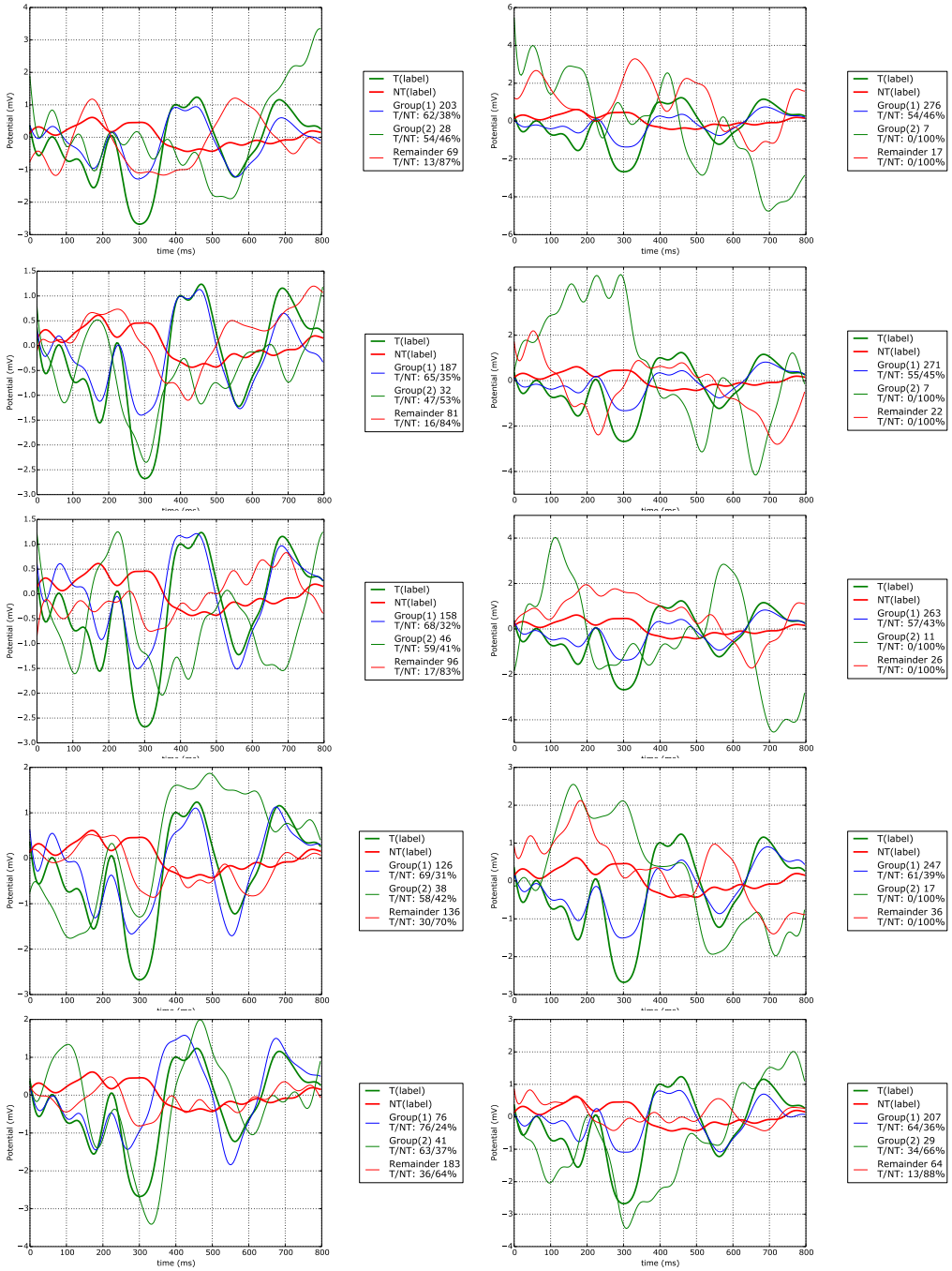

Figure B.20: Extracting two groups of epochs using the Kruskal-Walis test. On the left V, right dV. Showing results for participant 4

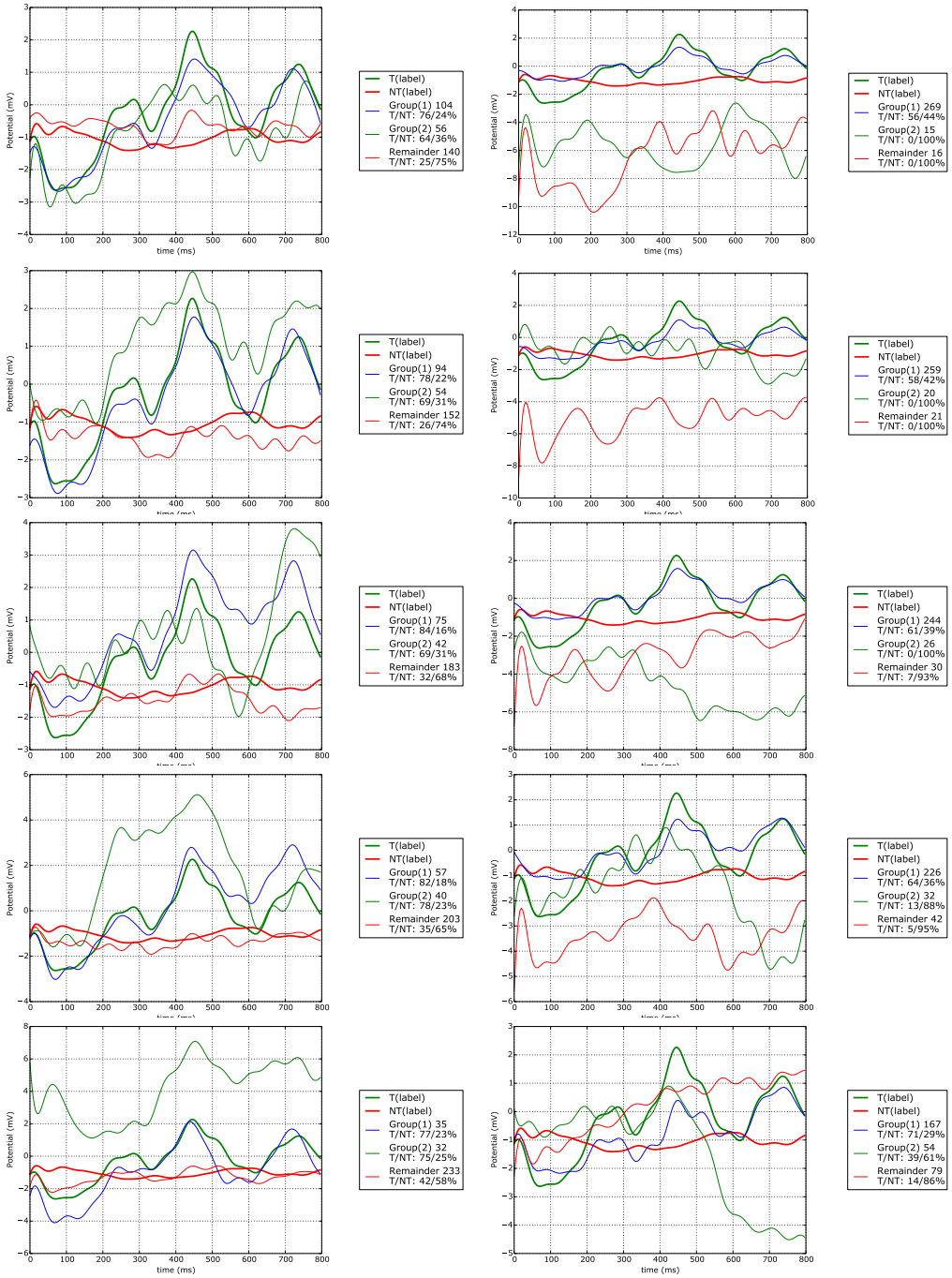

Figure B.21: Extracting two groups of epochs using the Kruskal-Walis test. On the left V, right dV. Showing results for participant 5

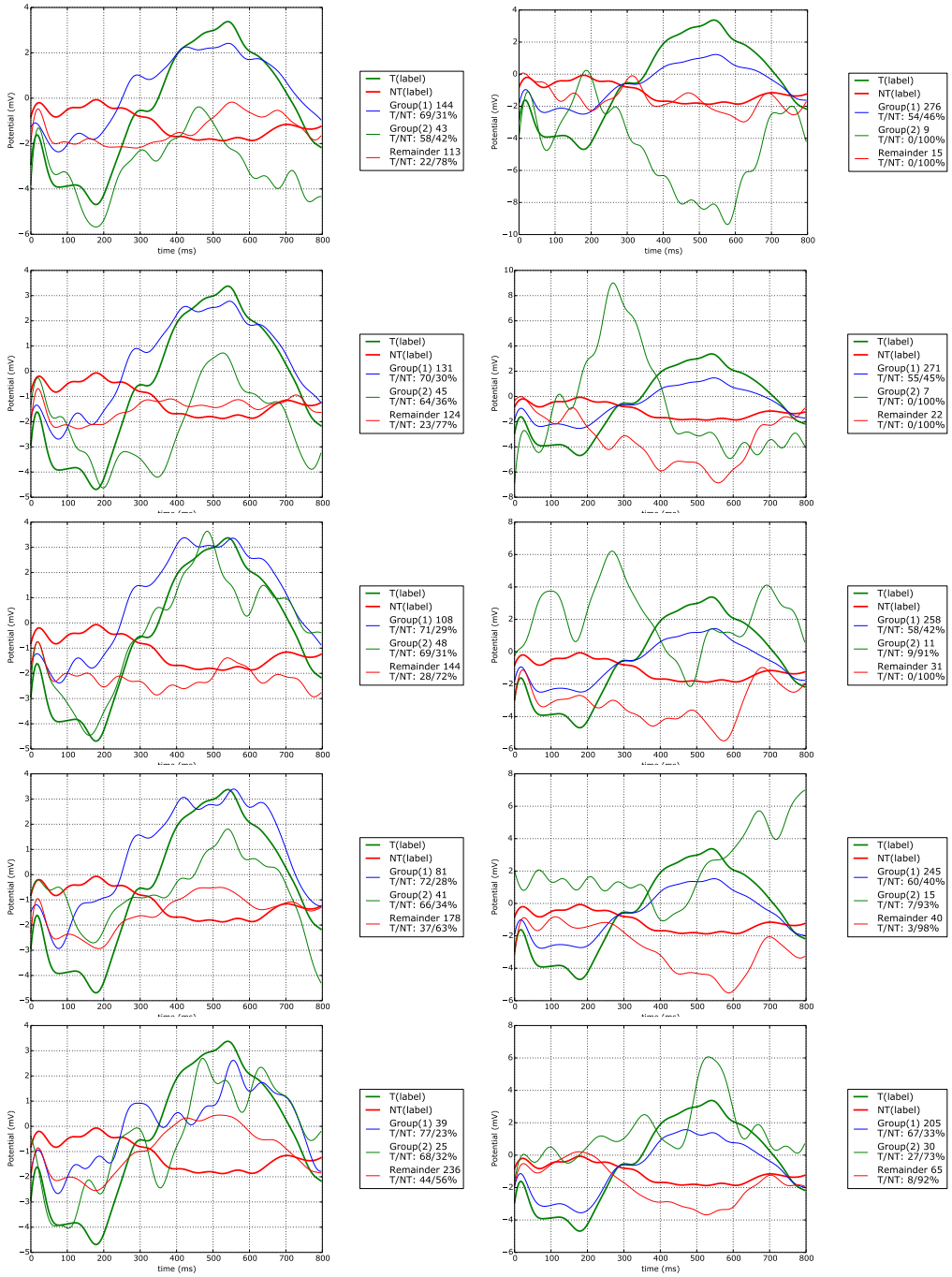

Figure B.22: Extracting two groups of epochs using the Kruskal-Walis test. On the left V, right dV. Showing results for participant 6

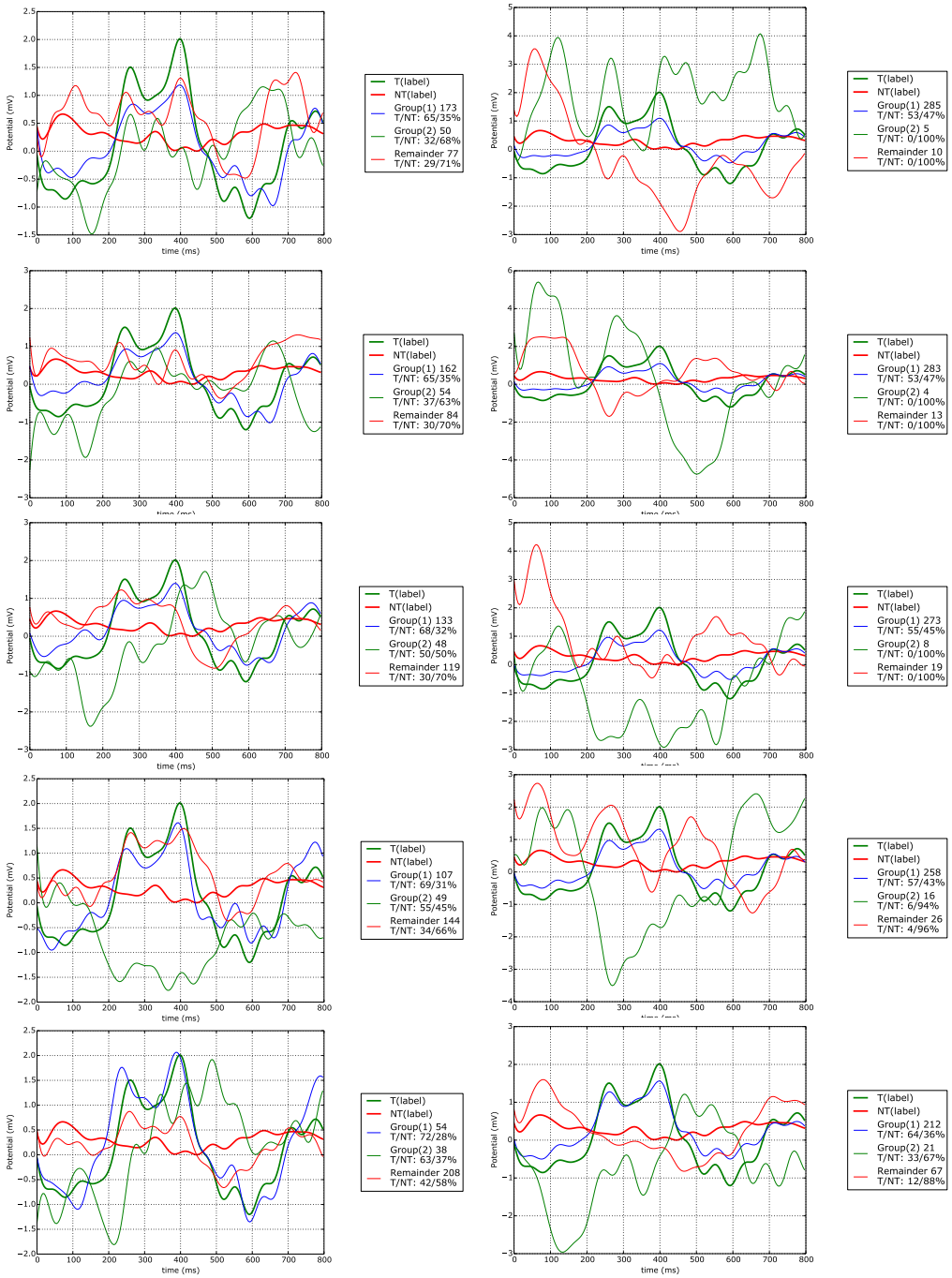

Figure B.23: Extracting two groups of epochs using the Kruskal-Walis test. On the left V, right dV. Showing results for participant 7

Figure B.24: Extracting two groups of epochs using the Kruskal-Walis test. On the left V, right dV. Showing results for participant 8

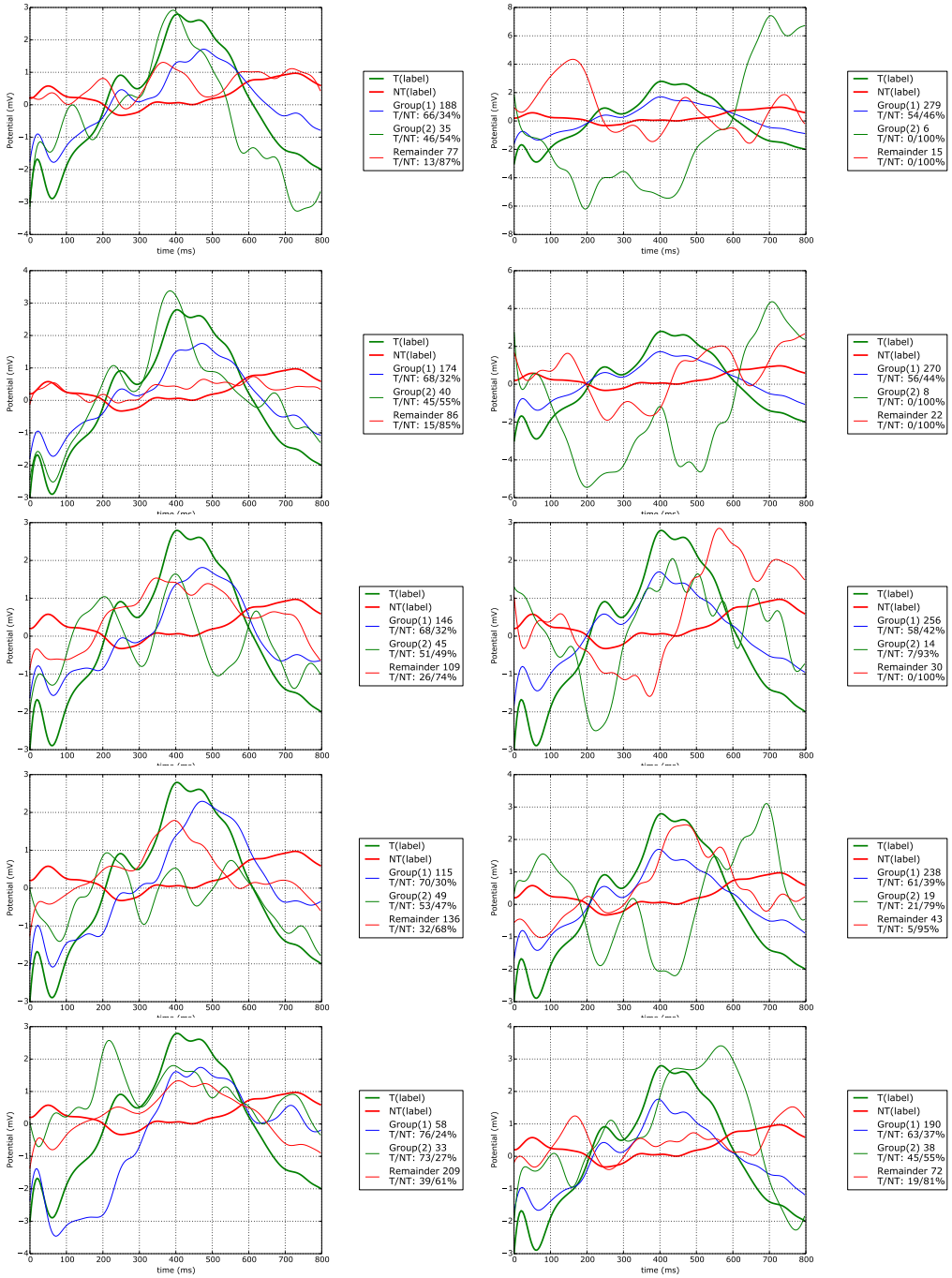

Figure B.25: Extracting two groups of epochs using the Kruskal-Walis test. On the left V, right dV. Showing results for participant 9

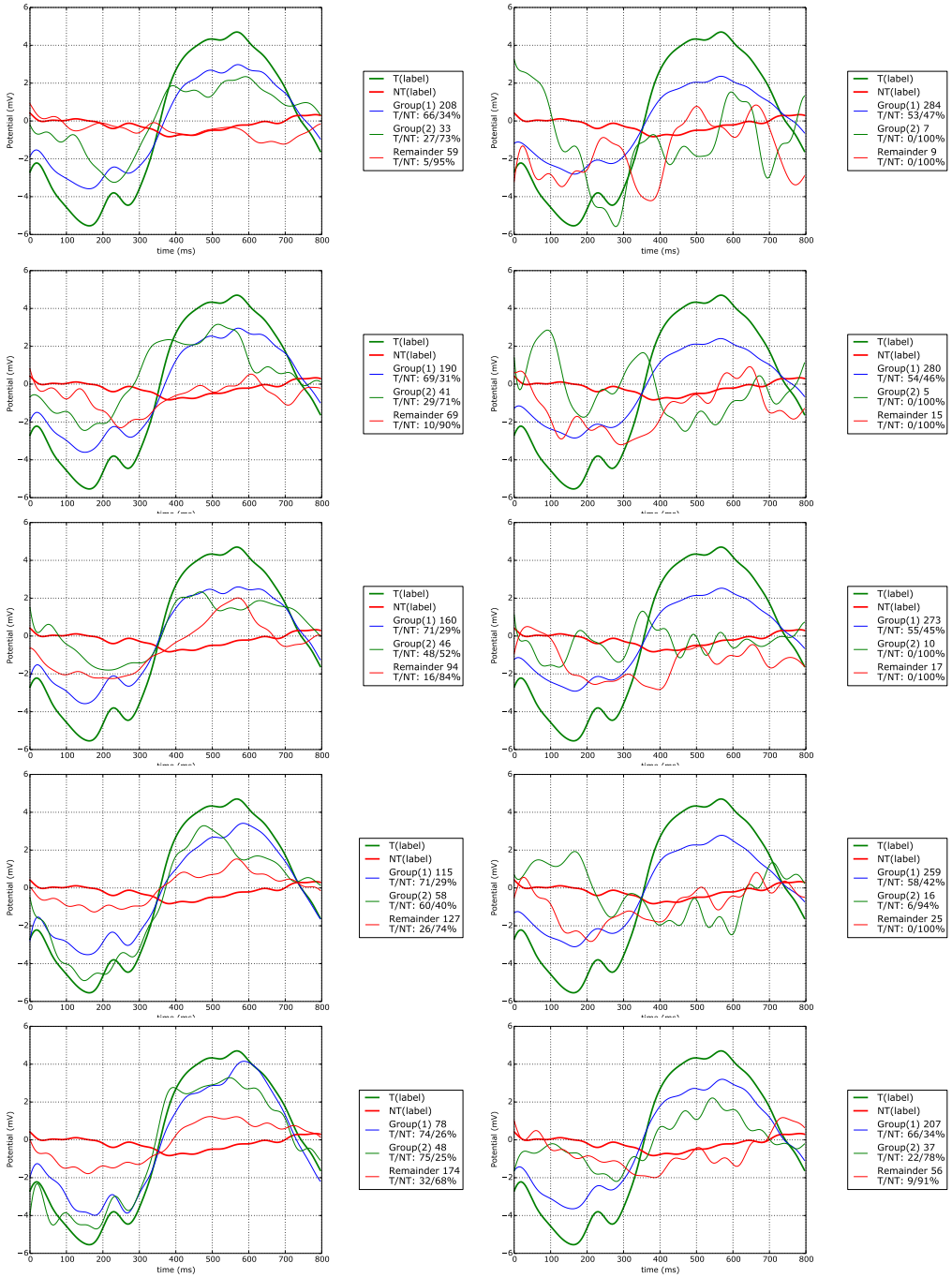

Figure B.26: Extracting two groups of epochs using the Kruskal-Walis test. On the left V, right dV. Showing results for participant 10

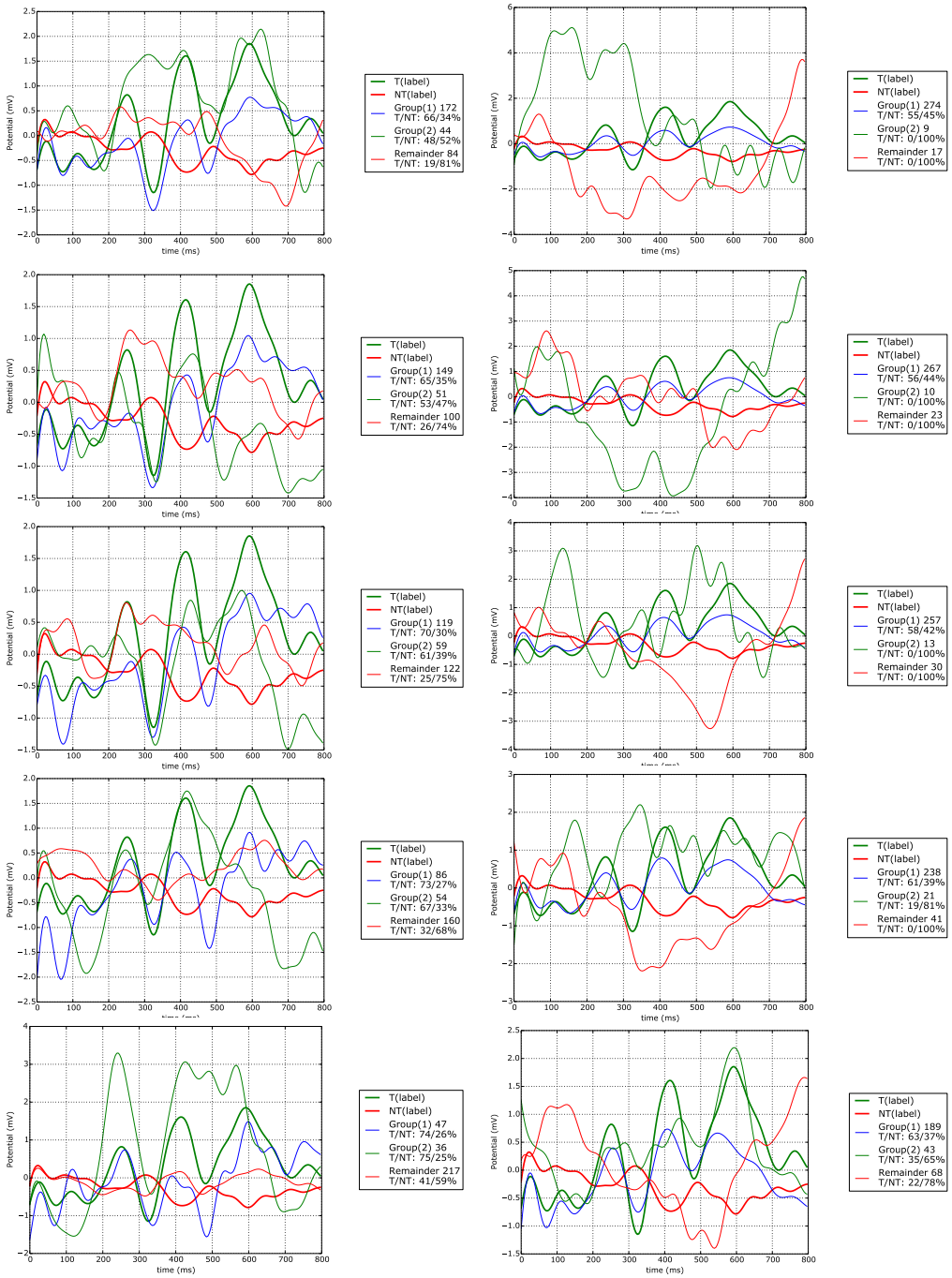

Figure B.27: Extracting two groups of epochs using the Kruskal-Walis test. On the left V, right dV. Showing results for participant 11

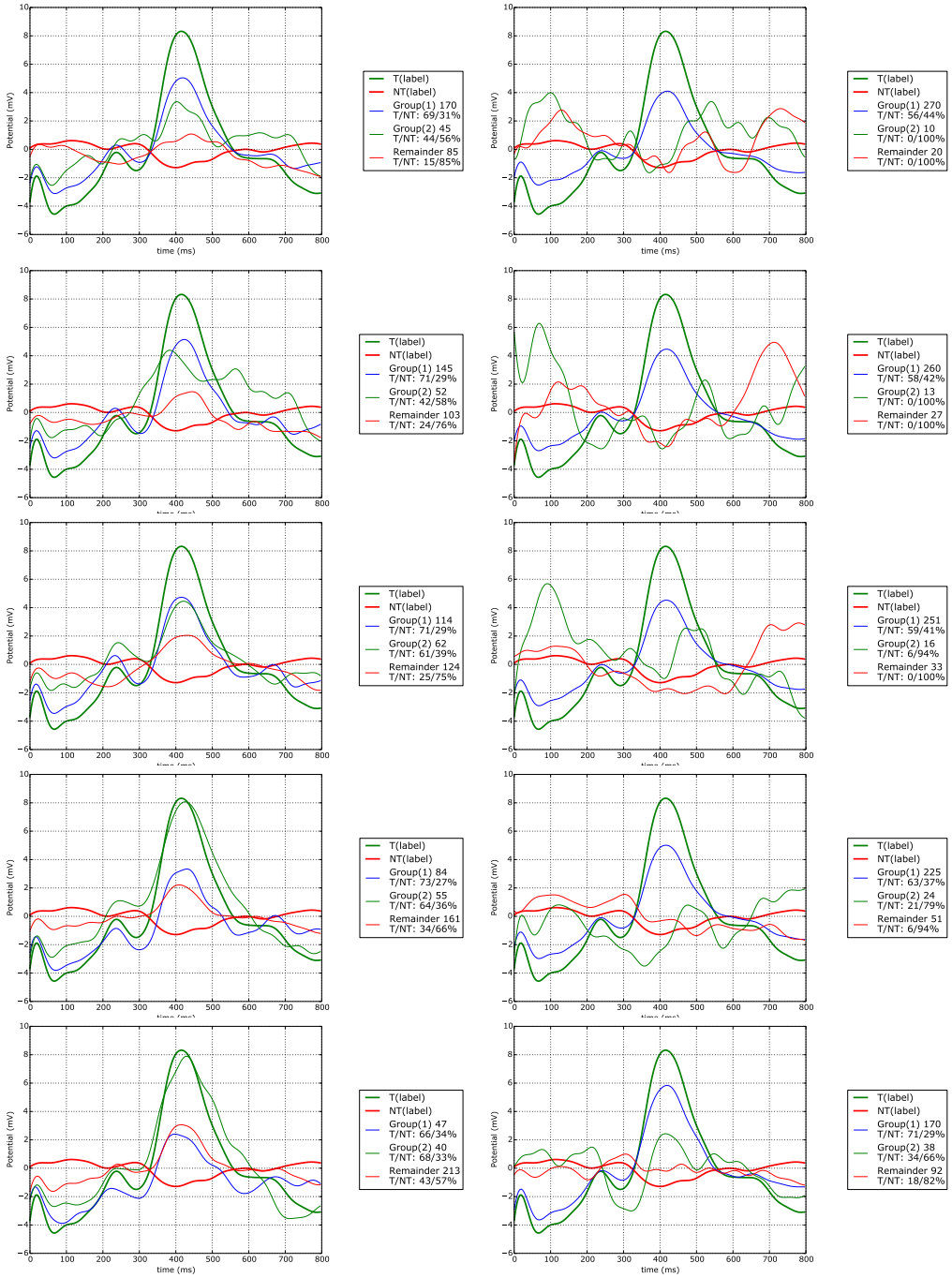

Figure B.28: Extracting two groups of epochs using the Kruskal-Walis test. On the left V, right dV. Showing results for participant 12

Figure B.29: Extracting two groups of epochs using the Kruskal-Walis test. On the left V, right dV. Showing results for participant 13

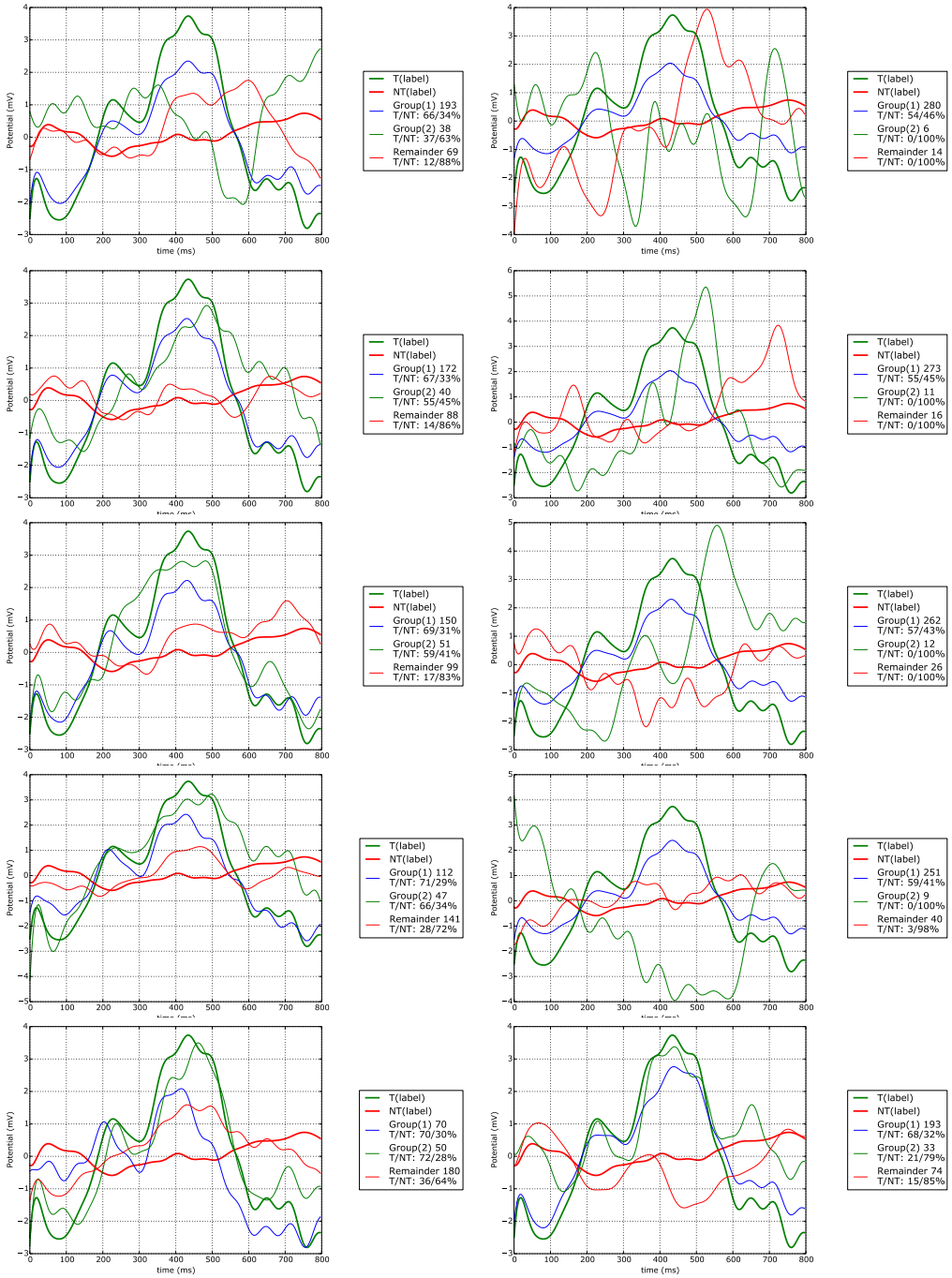

Figure B.30: Extracting two groups of epochs using the Kruskal-Walis test. On the left V, right dV. Showing results for participant 14

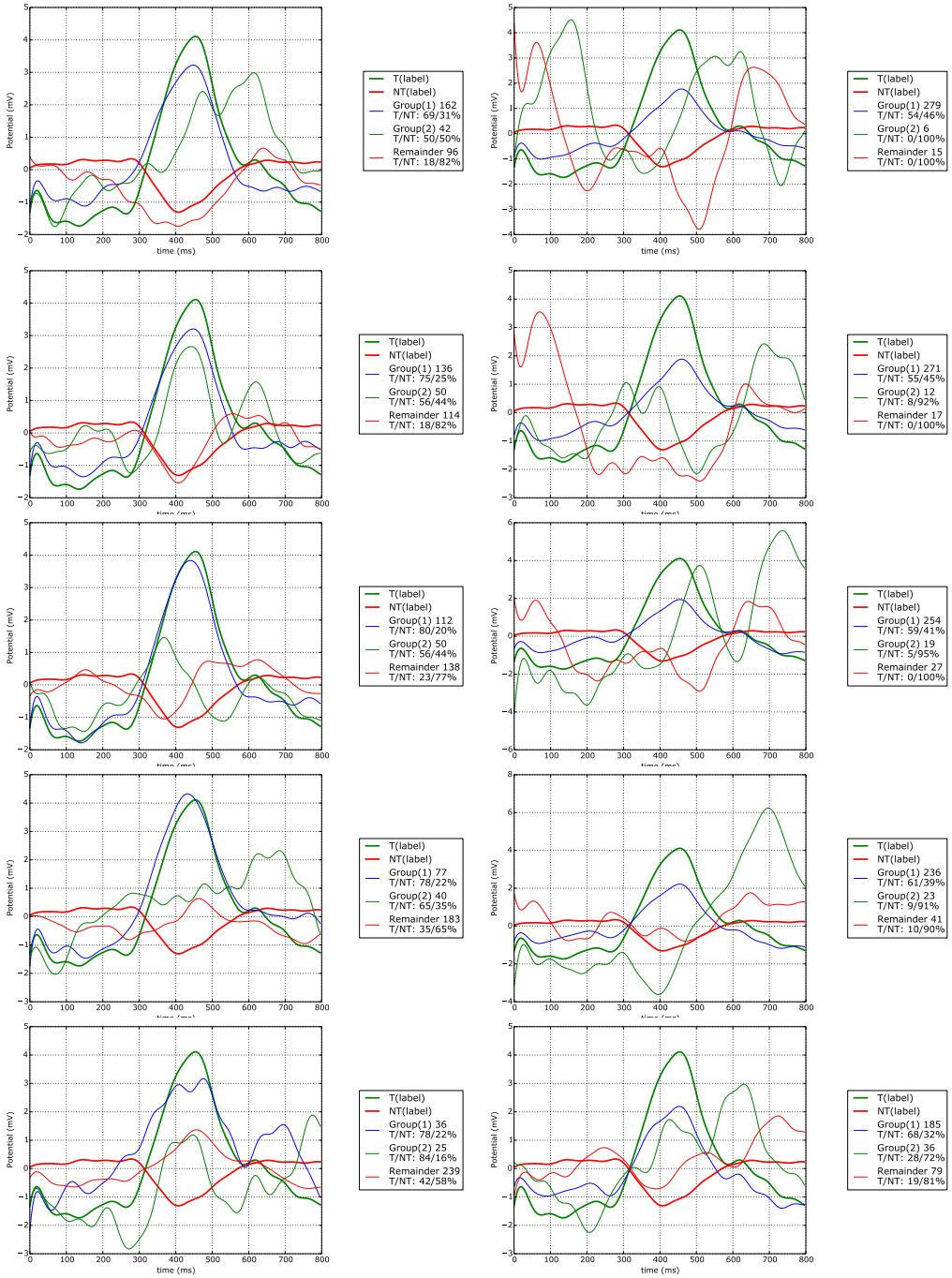

Figure B.31: Extracting two groups of epochs using the Kruskal-Walis test. On the left V, right dV. Showing results for participant 15

Figure B.32: Extracting two groups of epochs using the Kruskal-Walis test. On the left V, right dV. Showing results for participant 16

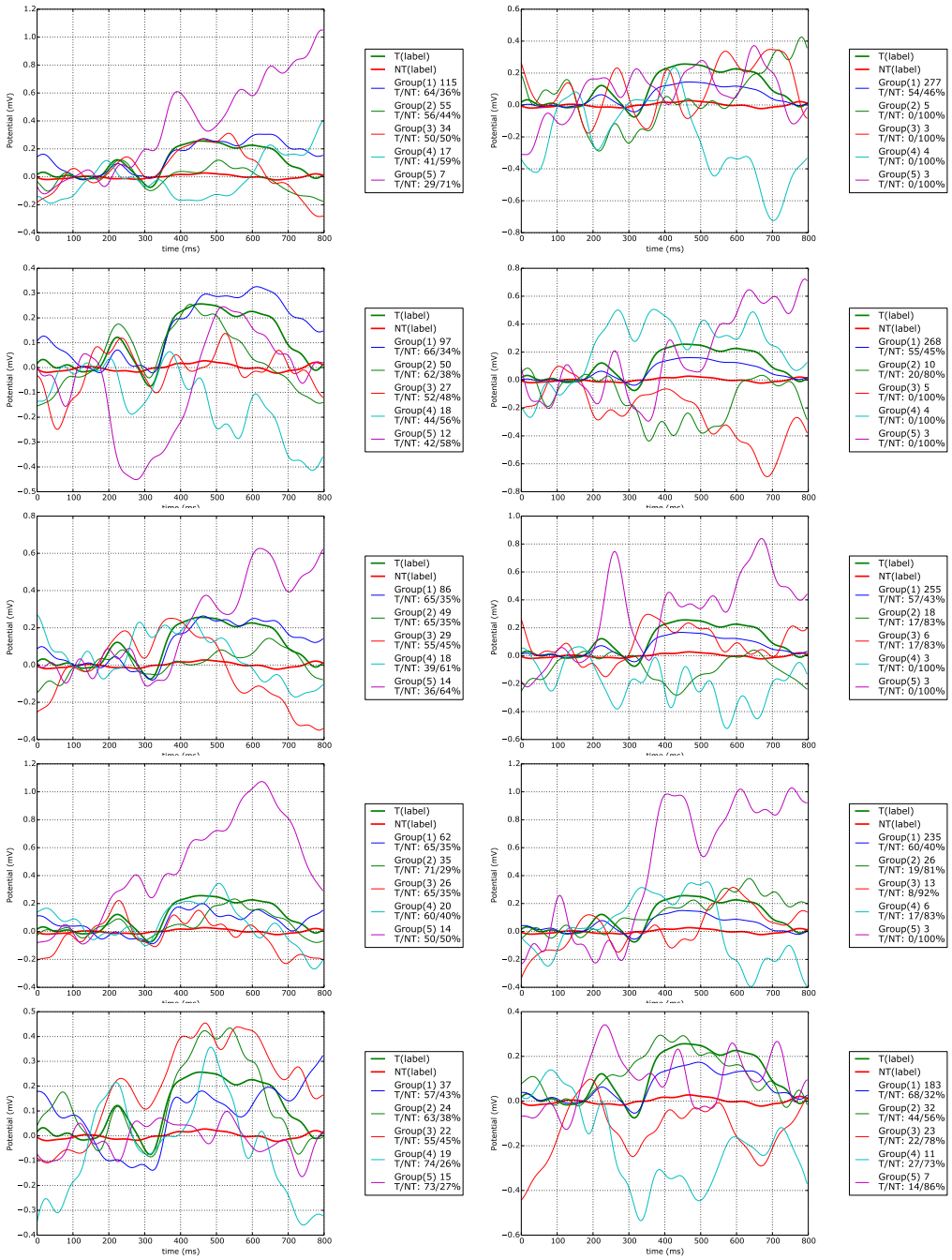

Figure B.33: Extracting 5 groups of epochs using the Kruskal-Walis test. On the left V, right dV. Showing results for participant 1

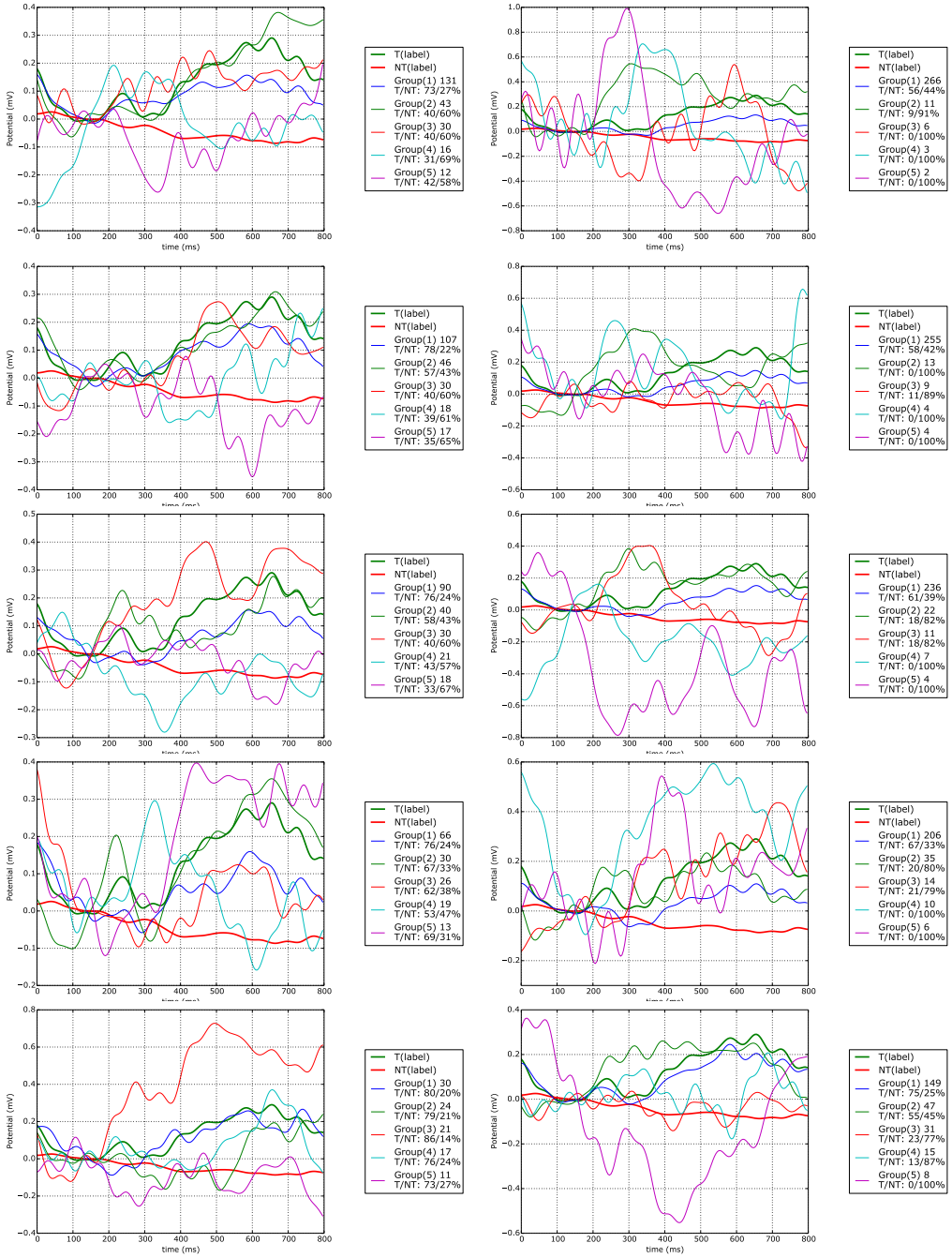

Figure B.34: Extracting 5 groups of epochs using the Kruskal-Walis test. On the left V, right dV. Showing results for participant 2

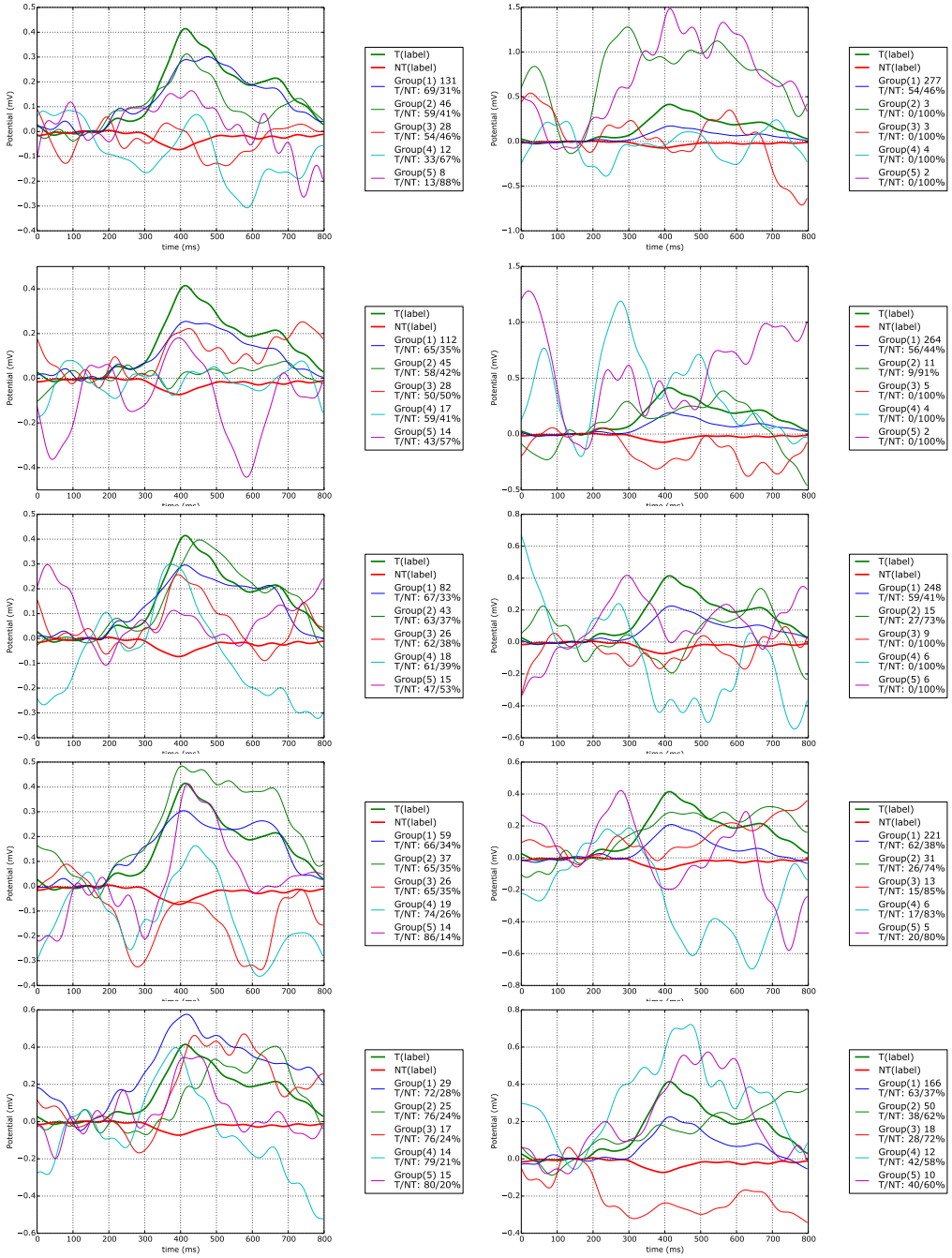

Figure B.35: Extracting 5 groups of epochs using the Kruskal-Walis test. On the left V, right dV. Showing results for participant 3

Figure B.36: Extracting 5 groups of epochs using the Kruskal-Walis test. On the left V, right dV. Showing results for participant 4

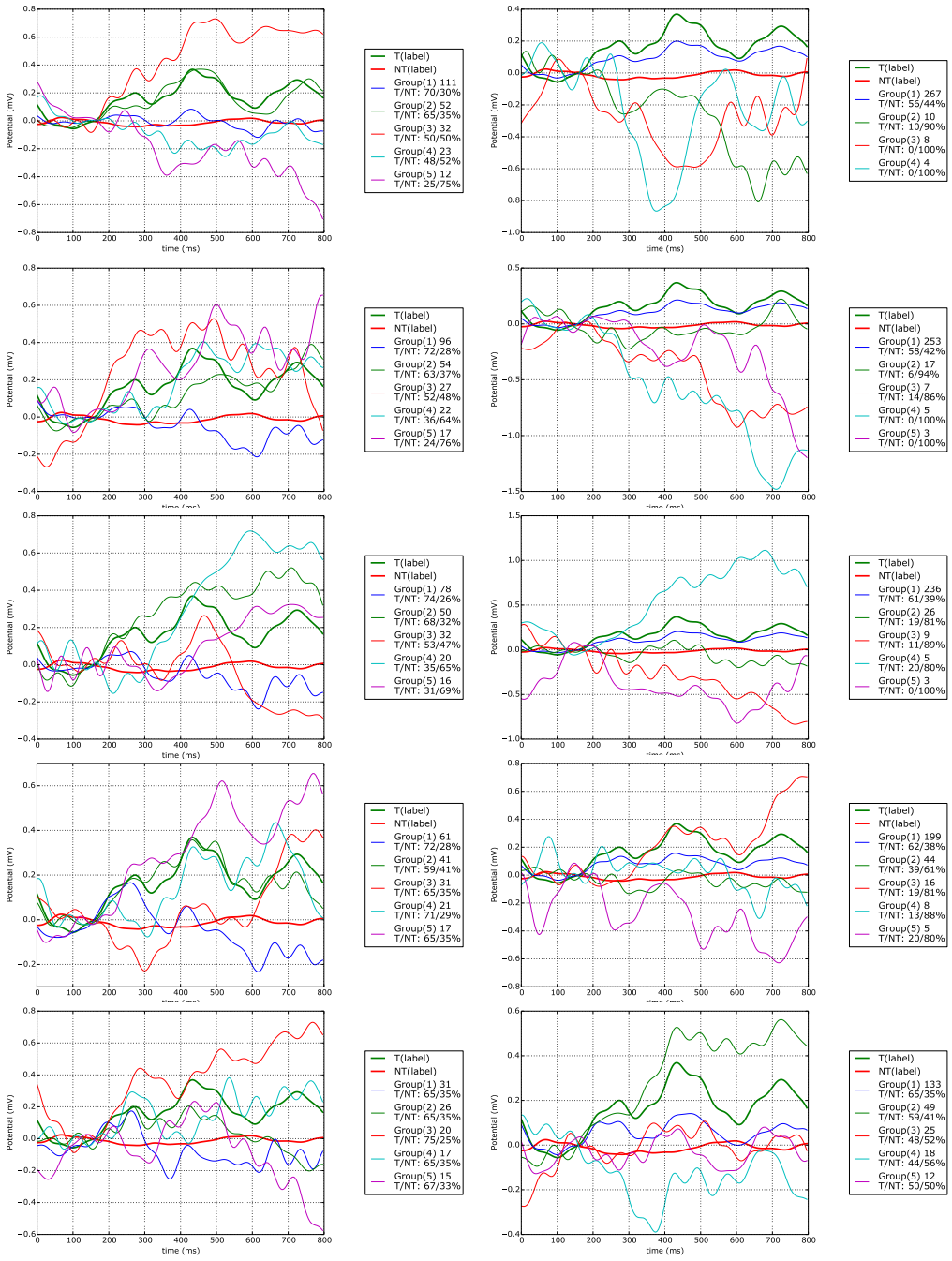

Figure B.37: Extracting 5 groups of epochs using the Kruskal-Walis test. On the left V, right dV. Showing results for participant 5

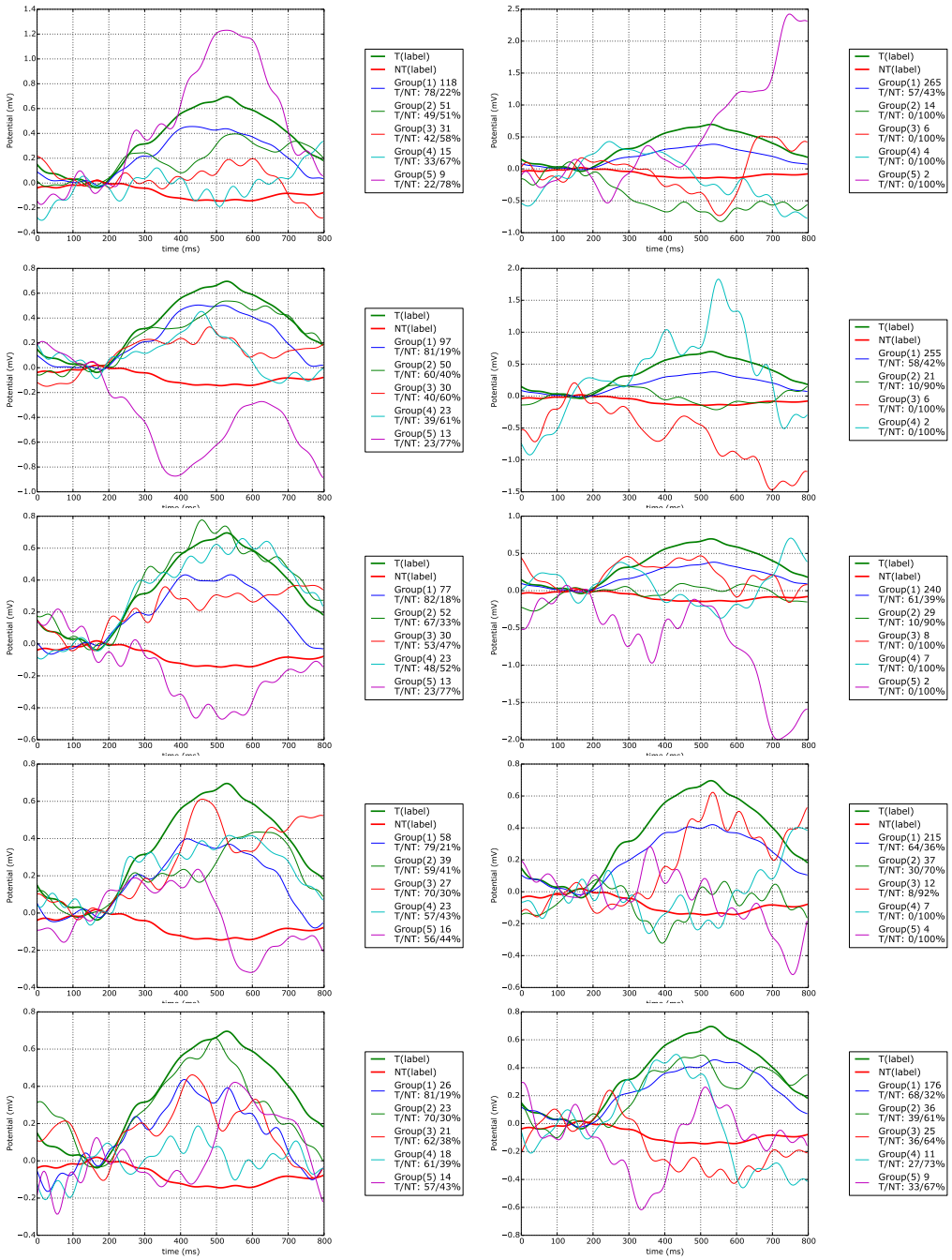

Figure B.38: Extracting 5 groups of epochs using the Kruskal-Walis test. On the left V, right dV. Showing results for participant 6

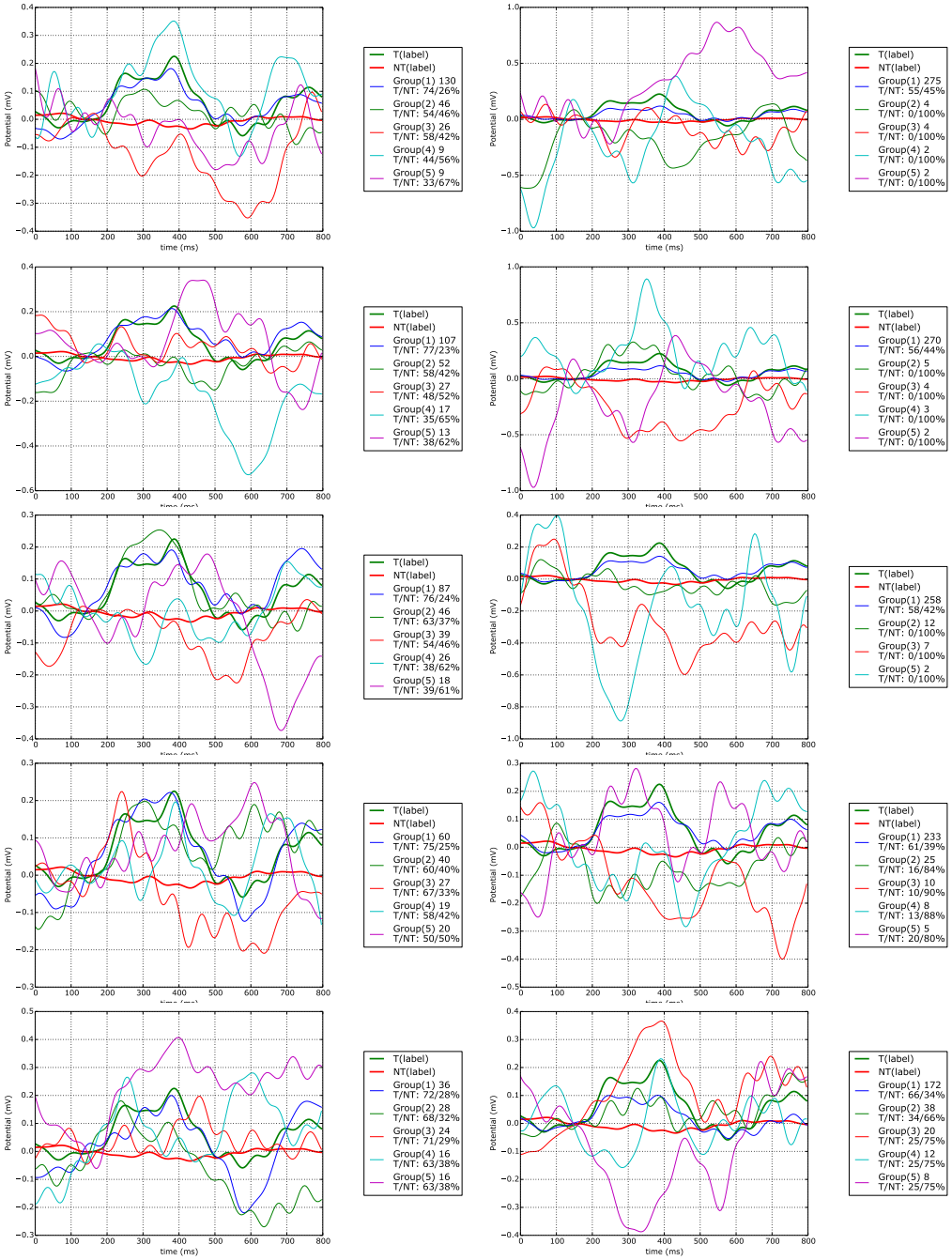

Figure B.39: Extracting 5 groups of epochs using the Kruskal-Walis test. On the left V, right dV. Showing results for participant 7

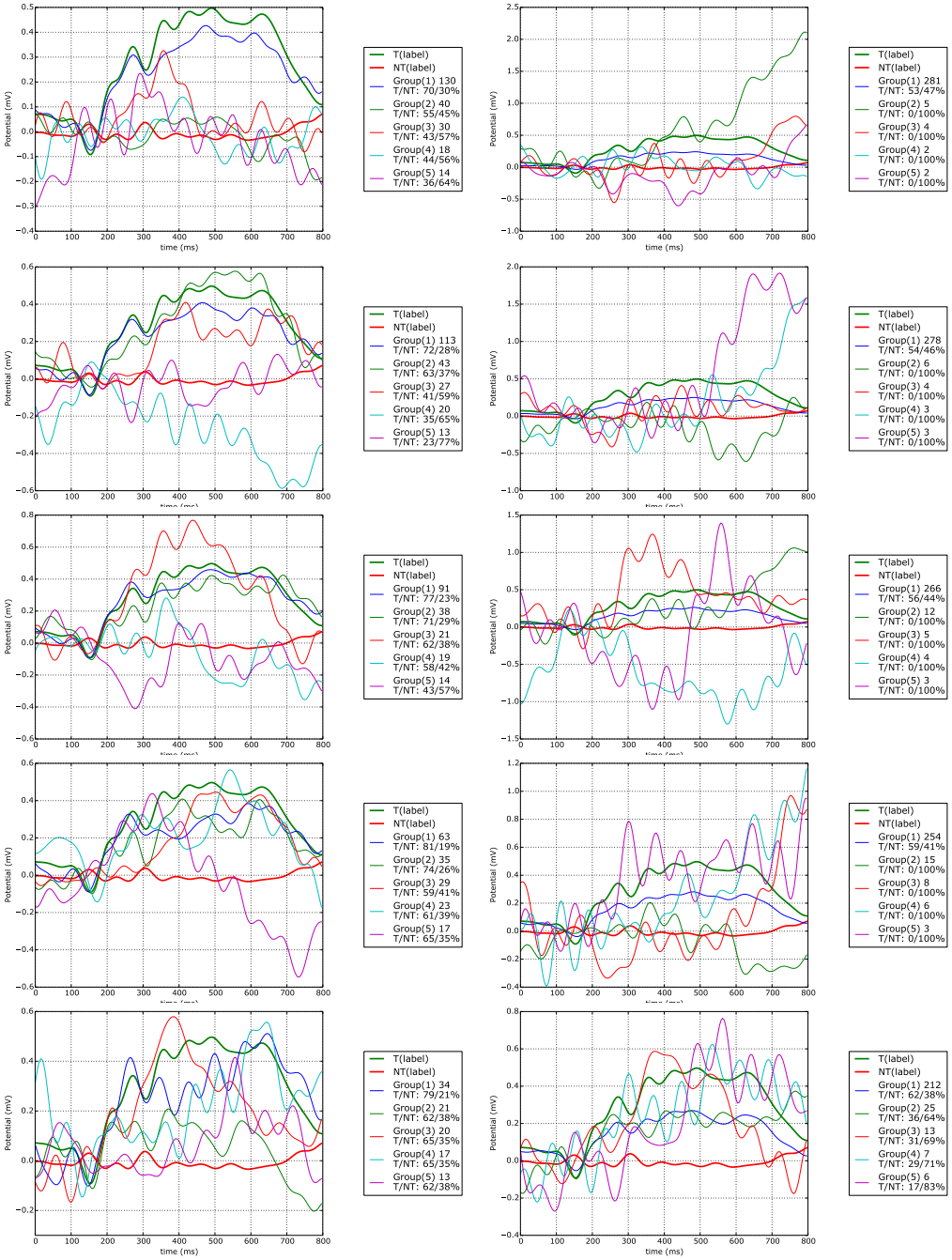

Figure B.40: Extracting 5 groups of epochs using the Kruskal-Walis test. On the left V, right dV. Showing results for participant 8

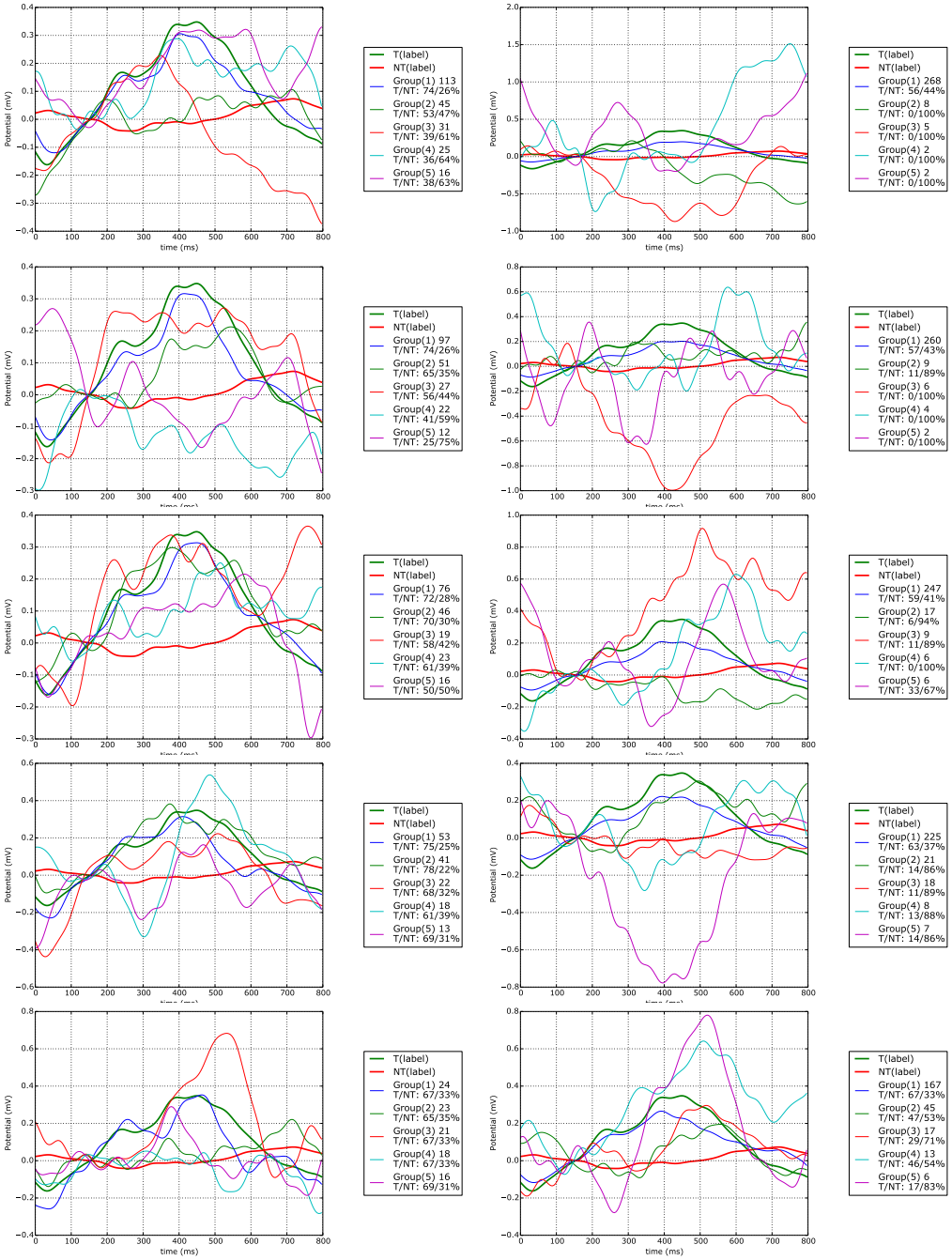

Figure B.41: Extracting 5 groups of epochs using the Kruskal-Walis test. On the left V, right dV. Showing results for participant 9

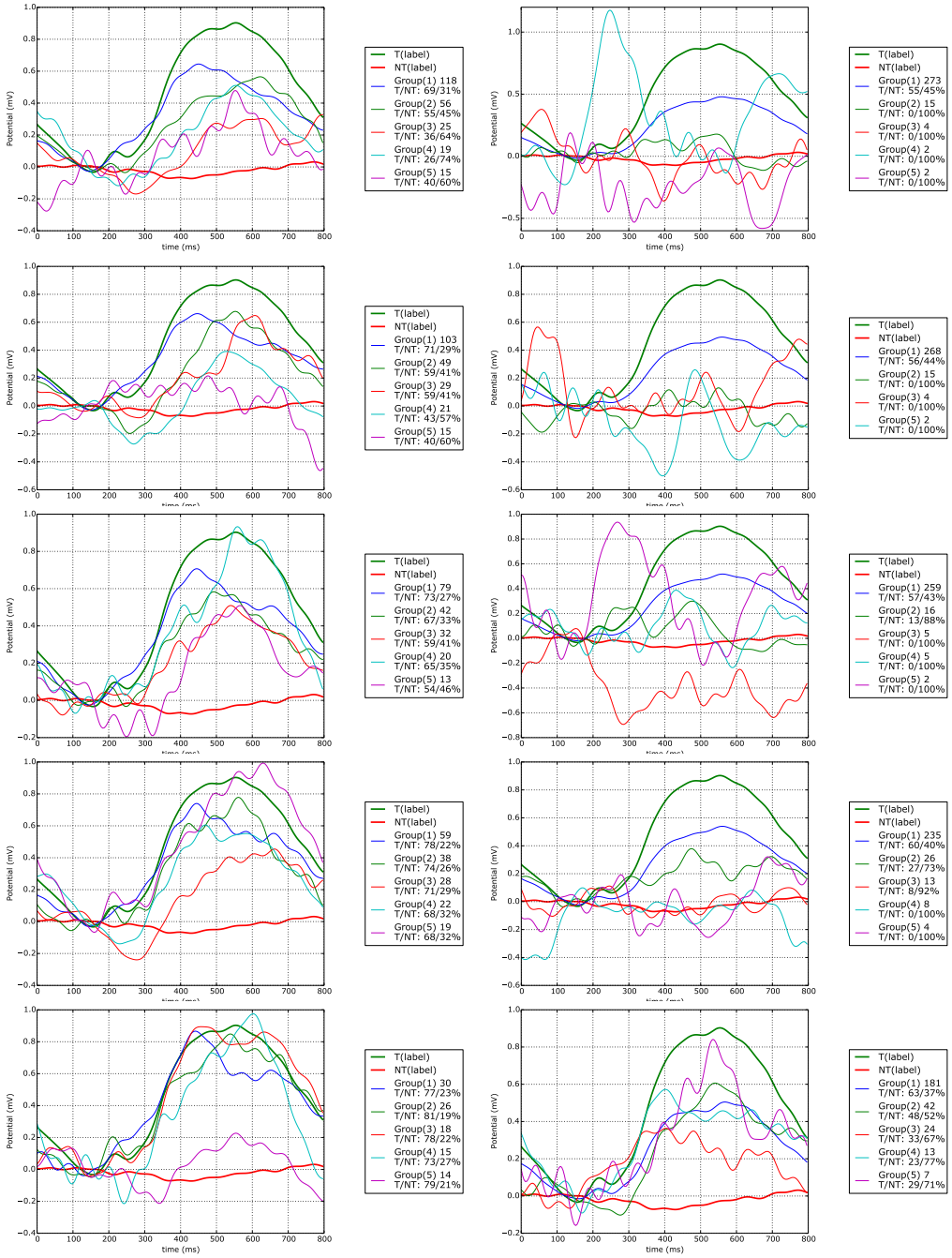

Figure B.42: Extracting 5 groups of epochs using the Kruskal-Walis test. On the left V, right dV. Showing results for participant 10

Figure B.43: Extracting 5 groups of epochs using the Kruskal-Walis test. On the left V, right dV. Showing results for participant 11

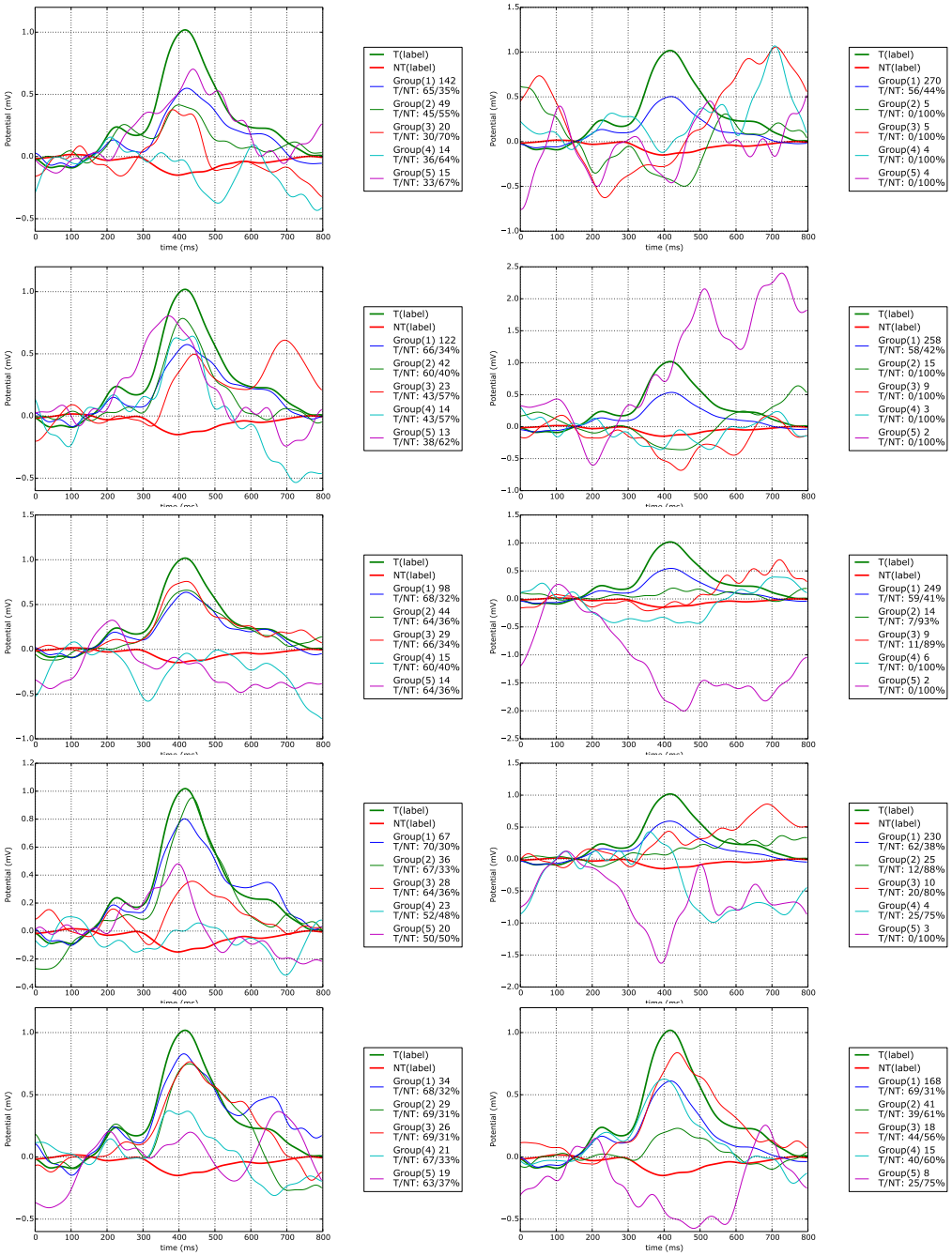

Figure B.44: Extracting 5 groups of epochs using the Kruskal-Walis test. On the left V, right dV. Showing results for participant 12

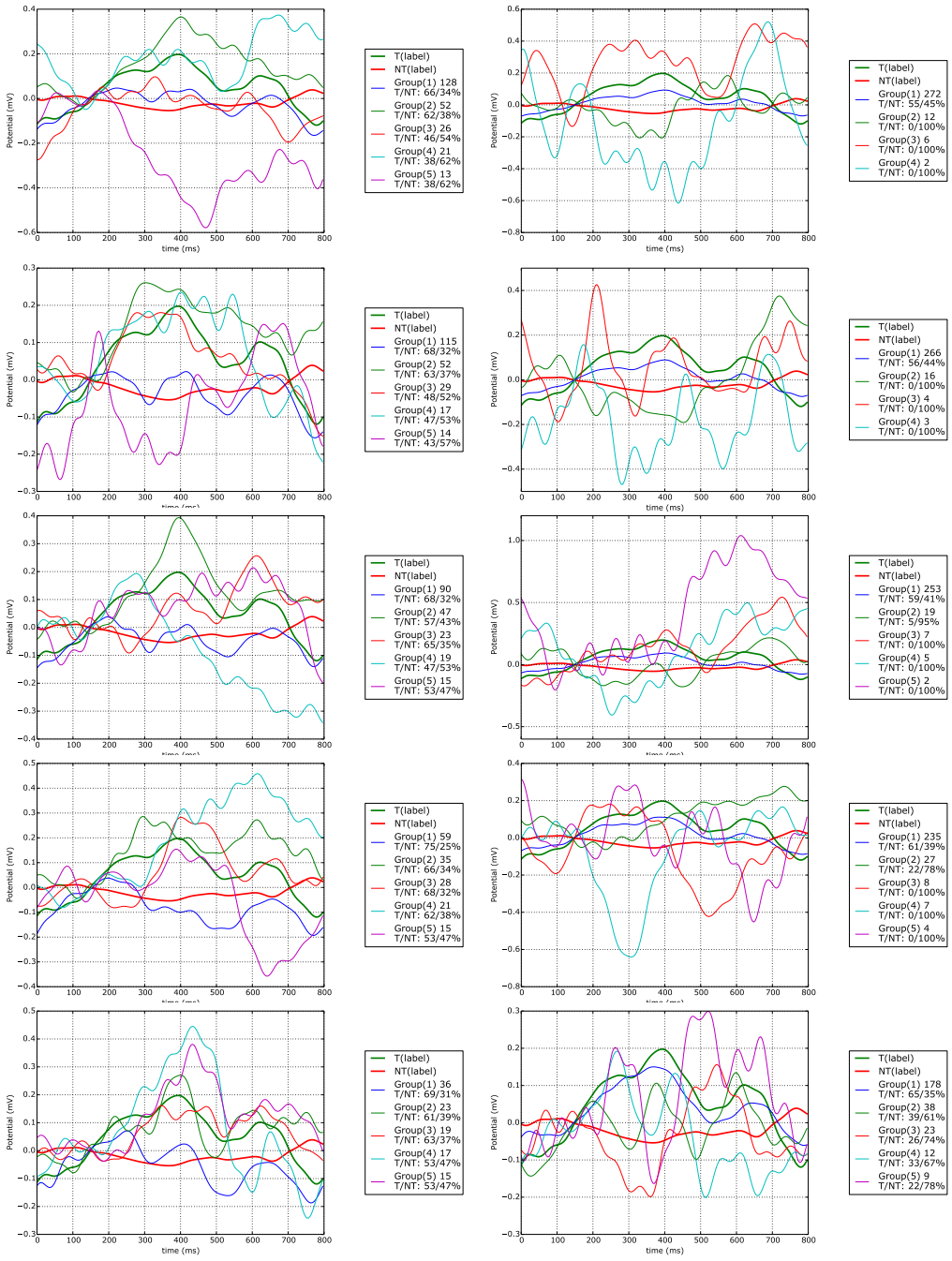

Figure B.45: Extracting 5 groups of epochs using the Kruskal-Walis test. On the left V, right dV. Showing results for participant 13

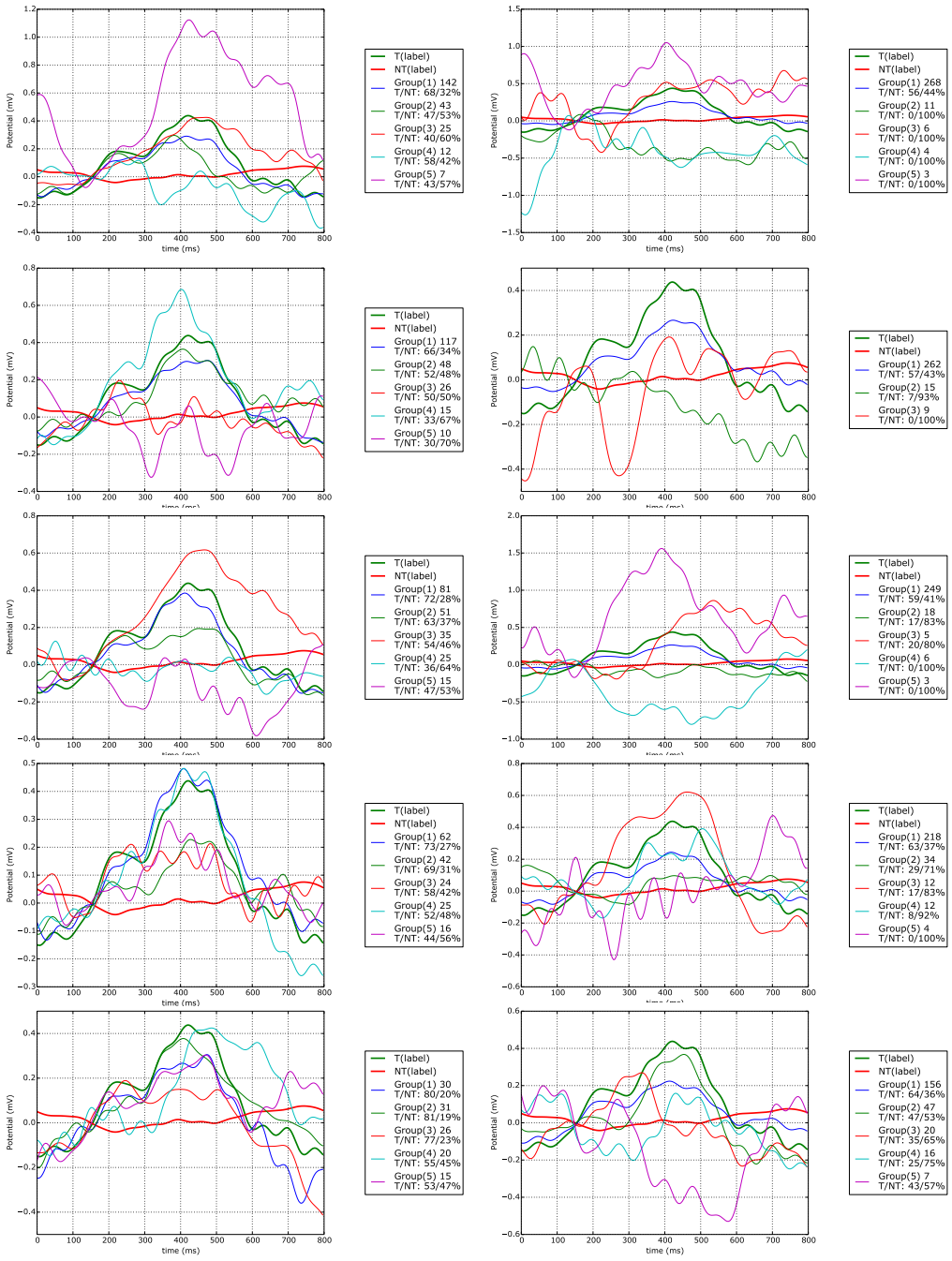

Figure B.46: Extracting 5 groups of epochs using the Kruskal-Walis test. On the left V, right dV. Showing results for participant 14

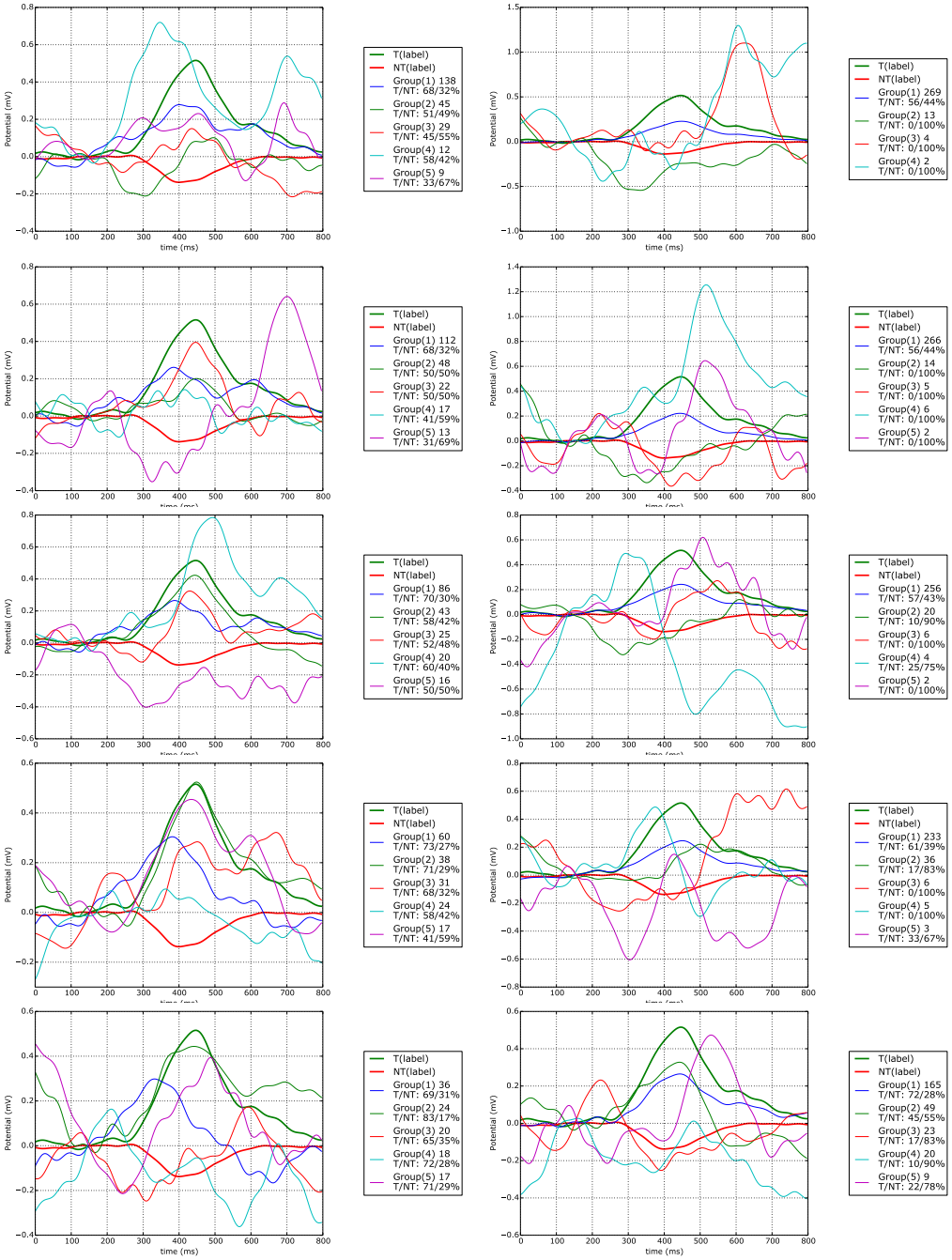

Figure B.47: Extracting 5 groups of epochs using the Kruskal-Walis test. On the left V, right dV. Showing results for participant 15

Figure B.48: Extracting 5 groups of epochs using the Kruskal-Walis test. On the left V, right dV. Showing results for participant 16