VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA INFORMAČNÍCH TECHNOLOGIÍ
ÚSTAV INFORMAČNÍCH SYSTÉMŮ

FACULTY OF INFORMATION TECHNOLOGY
DEPARTMENT OF INFORMATION SYSTEMS

TIME FREQUENCY ANALYSIS OF ERP SIGNALS

DIPLOMOVÁ PRÁCE
MASTER'S THESIS

AUTOR PRÁCE JAN BARTŮŠEK
AUTHOR

BRNO 2007

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA INFORMAČNÍCH TECHNOLOGIÍ
ÚSTAV POČÍTAČOVÉ GRAFIKY A MULTIMÉDIÍ

FACULTY OF INFORMATION TECHNOLOGY
DEPARTMENT OF COMPUTER GRAPHICS AND MULTIMEDIA

ČASOVĚ FREKVENČNÍ ANALÝZA ERP SIGNÁLŮ

TIME FREQUENCY ANALYSIS OF ERP SIGNALS

DIPLOMOVÁ PRÁCE
MASTER'S THESIS

AUTOR PRÁCE — JAN BARTŮŠEK
AUTHOR

VEDOUCÍ PRÁCE — DOC. DR. ING. JAN ČERNOCKÝ
SUPERVISOR

BRNO 2007

MSc. Thesis Specification/6121/2006/xbartu04

Brno University of Technology - Faculty of Information Technology

Department of Computer Graphics and Multimedia   Academic year 2006/2007

MSc. Thesis Specification

For: **Bartůšek Jan**

Branch of study: Computer Science and Engineering

Title: **Time Frequency Analysis of ERP Signals**

Category: Signály a systémy

Instructions for project work:

1. Get familiar with fundaments of signal processing and ICA approach. Get familiar with EEG and ERP fundaments and EEGLab toolbox for MATLAB.
2. Get familiar with the project on ERP features clustering supervised by P. Nikolopulos.
3. Choose one time-frequency technique, investigate its impact and benefit in context of improving the project.
4. Implement an algorithm demonstrating the capabilities of the selected technique.
5. Draw conclusions from your work and discuss how the results can be extended in the future.

Basic references:

- A. Hyvärinen and E. Oja. Independent Component Analysis: Algorithms and Applications. Neural Networks 13(4-5), pp. 411-430, 2000
- S. Makeig, A. Delorme, H. Serby: The EEGlab tutrial, Schwarz Center for Computational Neuroscience

The Term Project discussion items:

No requirements.

Detailed formal specifications can be found at http://www.fit.vutbr.cz/info/szz/

The MSc. Thesis must define its purpose, describe a current state of the art, introduce the theoretical and technical background relevant to the problems solved, and specify what parts have been used from earlier projects or have been taken over from other sources.

Each student will hand-in printed as well as electronic versions of the technical report, an electronic version of the complete program documentation, program source files, and a functional hardware prototype sample if desired. The information in electronic form will be stored on a standard medium (diskette, CD-ROM) in formats common at the FIT. In order to allow regular handling, the medium will be securely attached to the printed report.

Supervisor: **Černocký Jan, doc. Dr. Ing.**, DCGM FIT BUT

Beginning of work: November 1, 2006

Date of delivery: May 22, 2007

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
Fakulta informačních technologií
Ústav počítačové grafiky a multimédií
612 66 Brno, Božetěchova 2

L.S.

Pavel Zemčík

Associate Professor and Head of Department

LICENČNÍ SMLOUVA
POSKYTOVANÁ K VÝKONU PRÁVA UŽÍT ŠKOLNÍ DÍLO

uzavřená mezi smluvními stranami

1. Pan

Jméno a příjmení: **Jan Bartůšek**
Id studenta: 21941
Bytem: Měřičkova 52, 621 00 Brno
Narozen: 01. 04. 1982, Brno
(dále jen "autor")

a

2. Vysoké učení technické v Brně

Fakulta informačních technologií
se sídlem Božetěchova 2/1, 612 66 Brno, IČO 00216305
jejímž jménem jedná na základě písemného pověření děkanem fakulty:

....

(dále jen "nabyvatel")

Článek 1
Specifikace školního díla

1. Předmětem této smlouvy je vysokoškolská kvalifikační práce (VŠKP):
diplomová práce

Název VŠKP: Time Frequency Analysis of ERP Signals
Vedoucí/školitel VŠKP: Černocký Jan, doc. Dr. Ing.
Ústav: Ústav počítačové grafiky a multimédií
Datum obhajoby VŠKP: ....

VŠKP odevzdal autor nabyvateli v:

tištěné formě počet exemplářů: 1
elektronické formě počet exemplářů: 2 (1 ve skladu dokumentů, 1 na CD)

2. Autor prohlašuje, že vytvořil samostatnou vlastní tvůrčí činností dílo shora popsané a specifikované. Autor dále prohlašuje, že při zpracovávání díla se sám nedostal do rozporu s autorským zákonem a předpisy souvisejícími a že je dílo dílem původním.
3. Dílo je chráněno jako dílo dle autorského zákona v platném znění.
4. Autor potvrzuje, že listinná a elektronická verze díla je identická.

Článek 2
Udělení licenčního oprávnění

1. Autor touto smlouvou poskytuje nabyvateli oprávnění (licenci) k výkonu práva uvedené dílo nevýdělečně užít, archivovat a zpřístupnit ke studijním, výukovým a výzkumným účelům včetně pořizování výpisů, opisů a rozmnoženin.
2. Licence je poskytována celosvětově, pro celou dobu trvání autorských a majetkových práv k dílu.
3. Autor souhlasí se zveřejněním díla v databázi přístupné v mezinárodní síti:
   - ☐ ihned po uzavření této smlouvy
   - ☐ 1 rok po uzavření této smlouvy
   - ☐ 3 roky po uzavření této smlouvy
   - ☐ 5 let po uzavření této smlouvy
   - ☐ 10 let po uzavření této smlouvy

   (z důvodu utajení v něm obsažených informací)
4. Nevýdělečné zveřejňování díla nabyvatelem v souladu s ustanovením § 47b zákona č. 111/ 1998 Sb., v platném znění, nevyžaduje licenci a nabyvatel je k němu povinen a oprávněn ze zákona.

Článek 3
Závěrečná ustanovení

1. Smlouva je sepsána ve třech vyhotoveních s platností originálu, přičemž po jednom vyhotovení obdrží autor a nabyvatel, další vyhotovení je vloženo do VŠKP.
2. Vztahy mezi smluvními stranami vzniklé a neupravené touto smlouvou se řídí autorským zákonem, občanským zákoníkem, vysokoškolským zákonem, zákonem o archivnictví, v platném znění a popř. dalšími právními předpisy.
3. Licenční smlouva byla uzavřena na základě svobodné a pravé vůle smluvních stran, s plným porozuměním jejímu textu i důsledkům, nikoliv v tísni a za nápadně nevýhodných podmínek.
4. Licenční smlouva nabývá platnosti a účinnosti dnem jejího podpisu oběma smluvními stranami.

V Brně dne: ....

....	....
Nabyvatel	Autor

Abstrakt

Tato práce se zabývá vylepšením algoritmu pro sdružování (clustering) ERP signálů pomocí analýzy časových a prostorových vlastností pseudo-signálů získaných za pomocí metody analýzy nezávislých komponent (Independent Component Analysis). Naším zájmem je nalezení nových vlastností, které by zlepšily stávající výsledky. Tato práce se zabývá použitím Fourierovy transformace (Fourier Transform), FIR filtru a krátkodobé Fourierovy transformace ke zkvalitnění informace pro sdružovací algoritmy. Princip a použitelnost metody jsou popsány a demonstrovány ukázkovým algoritmem. Výsledky ukázaly, že pomocí dané metody je možné získat ze vstupních dat zajímavé informace, které mohou být úspěšně použity ke zlepšení výsledků. Výsledky nicméně nebyly konzultovány s neurology a úspěšnost metody byla posouzena pouze na základě srovnání ukázkového algoritmu s původní metodou.

Klíčová slova

časově frekvenční analýza, Analýza nezávislých komponent, zpracování signálů, Fourierova transformace , krátkodobá Fourierova transformace, elektroencefalograf, EEG, ERP, EP, FFT, STFT, ICA

Abstract

The aim of this work is to improve the algorithm for clustering ERP signals based on the temporal and spatial properties of pseudo-signals gained by the Independent Component Analysis. The main purpose is to find new features, which could improve the original algorithm. This technical report is investigating application of new features gained by Fourier Transform and Short Time Fourier Transform methods. Basic principle and performance of the concept is demonstrated on the sample algorithm. Although the results have not been consulted with neurospecialists, the comparison with results of the original algorithm has shown that the method can bring an interesting contribution to the original project and should be considered as a convenient improvement.

Keywords

time frequency analysis, independent component analysis, signal processing, Fourier transform, Short time Fourier transform, electroencephalogram, EEG, ERP, EP, FFT, STFT, ICA

Citace

Bartůšek Jan: Time Frequency Analysis of ERP signals. Brno, 2007, diplomová práce, FIT VUT v Brně.

Time frequency analysis of ERP signals

Prohlášení

Prohlašuji, že jsem tuto diplomovou práci vypracoval samostatně pod vedením Doc. Dr. Ing. Jana Černockého a prof. George Papadourakise (TEI of Crete, Řecko)

Další informace mi poskytl Panagiotis Nikolopulos (TEI of Crete, Řecko)

Uvedl jsem všechny literární prameny a publikace, ze kterých jsem čerpal.

Jan B.

....
Jméno Příjmení
Datum

12.5.2007

Content

1 Introduction

As a student of Brno University of Technology, Czech Republic, I got a chance to elaborate my final project work at Technological and Educational Institute of Crete (TEI) in Heraklion, Greece. Its main intent is to explore the time frequency properties of electroencephalogram (EEG) signals, more exactly Event Related Potentials (ERP), in the context of the main project work, which was simultaneously evolved in the same laboratory - Brain Special Interest Group of TEI Crete (BSIG).

This project work was originally meant as a reference investigation for co-workers of a BSIG lab, who design an application for separation and grouping of ERP signals and have not yet deled with the approach of time frequency analysis. The context of these two projects is closely interconnected, which is reflected in whole the work. The first chapters are presenting the basic fundaments and techniques of the original project, which is referred in this work as a Source Localization Algorithm (SLA). The necessity of time frequency analysis is concluded than and the presentation of its strengths and possible advantages is explained and demonstrated on proposed sample algorithm. Finally, the methods are discussed and the conclusion is given with proposal of the concept integration into the SLA solution.

Although the time frequency analyses are one of the most frequently used techniques in the EEG analysis, it still has not been considered as a successful approach for extracting features for grouping of ERP signals in SLA approach. The intent was not to give a description and evaluation of all or high amount of time frequency analysis approaches. The favorable way was rather to pick one or two methods and carefully analyze them. Three techniques are presented and discussed in the study: Use of Fast Fourier Transform for artifact rejection, classification of statistics of filtered signals and feature extraction based on the Short Time Fourier Transform approach. All the methods are applied on a data preprocessed by Independent Component Analysis (ICA) algorithm, a modern approach for hidden information extraction from linear signal mixtures. As the FFT transform approach has not shown to be a successful, the filtered signal statistics classification and STFT analysis proved to be interesting alternatives that require consideration of their implementation in corresponding SLA algorithm. The proposed algorithms are implemented in the MATLAB framework and their results are presented and described in the work.

Along with an introduction and conclusion parts, the work has two important sections describing the background and proposed solution. The background part introduces the EEG concept and the basic techniques used for its processing in the context of this work. The basic explanation of the original SLA algorithm and ICA method are following. At the end, after introduction of Fourier Transform and STFT a brief history of EEG time frequency analysis is given along with its strengths and reasons, why to use them in the SLA approach.

The second main section with proposed solution explains the possible time frequency approaches, possibilities of their use and their performance. The appendix part contains brief implementation description and more detailed program outputs.

This work neither extends nor arises from Term project work or Year project work of the study program of my home university.

2 The background

2.1 Electroencephalogram and ERP

2.1.1 Technical principles and basic descriptors

Electroencephalography is the oldest non-invasive method in investigating neural activity. Since the first time it was measured on humans in early 19th century, it has developed to the widely used and essential method of neuroscience.

The electroencephalograph can be roughly described as the mean value of the electrical activity (the sum of currents) of larger groups of neurons on the surface of the scalp. The electrodes of device can measure the potential difference in the relation to each other or with in-active reference electrode. These pairs are than interconnected to amplifiers and filtered by high-pass and low-pass filters. Recordings are usually projected on the paper or the screen of computer. The most common scheme defining the organization of electrodes is called "10-20 system" [11]. The label of every electrode consists of a letter, which describes the area on the scalp, and a number, which determines its more exact position. For more details, see Figure 2.1.

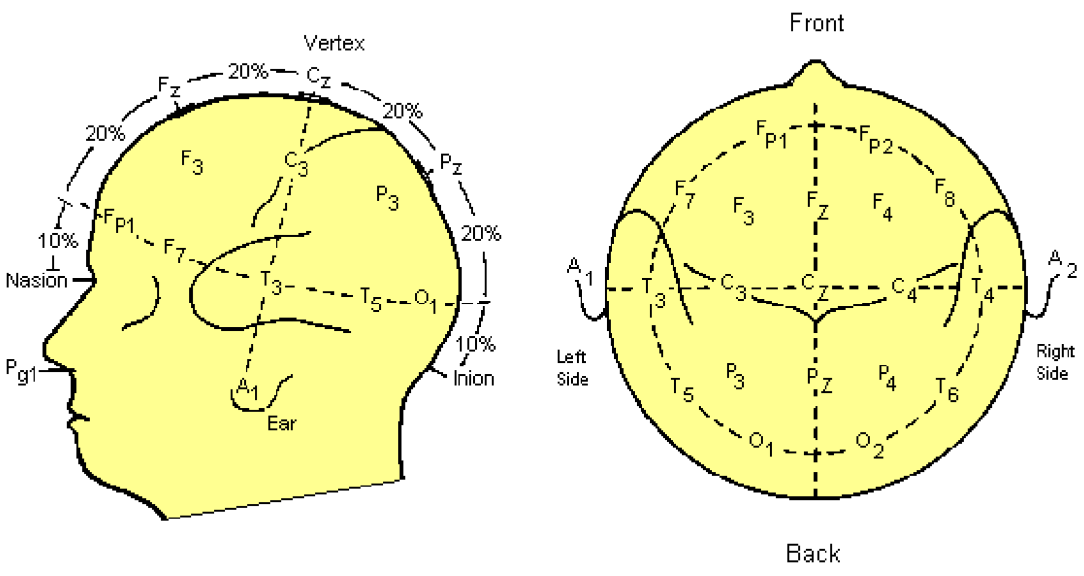

Figure 2.1 10-20 electrode placement system of EEG. The letters T, C, F, P, O and A stands for Temporal, Central, Frontal, Parietal, Occipital and Earlobe areas. Even numbers refer to the right hemisphere and odd numbers to the left one. Smaller numbers are closer to the central axis.

Many descriptors used for analysis of EEG signals are analyzed in the literature. Frequency, amplitude (maximum variance), shape, distribution, symmetry, synchrony, rythmicity and periodicity are some of the most widely used terms for description of EEG signals. The frequency analysis is the most common quantitative method for investigating EEG signals [12]. When Hans Berger was publishing his results of first measuring of human EEG signals in history [13], he already mentioned a presence of alpha and beta frequency oscillations. The study of different rhythmicities and their relations with state of mind has been keeping the attention of scientists until present days. After an intensive research, the frequency characteristics of EEG signal were divided into different classes and interconnected with typical patterns of brain's behavior, function and state [14]. The slowest Delta rhythms are components of frequency 0.5 Hz – 3.5 Hz, Theta rhythms denote frequency range 3.5 Hz – 7.5 Hz, Alpha rhythms 7.5 Hz – 12.5 Hz, Beta rhythms 12.5 – 30 Hz and Gamma rhythms 30 Hz – 60 Hz. The importance of frequency analysis is highlighted by current clinical use, where it is common approach used e.g. for epilepsy diagnostics.

2.1.2 Artifacts

The most problematic issue connected with investigation of EEG that is troubling the researchers are artifacts. Artifacts are all the components present in the output of EEG, which are not recordings of the brain activity. A great variety of different artifact types is described in the literature. Artifacts can be divided into three basic groups according to their origin (source): from the patient, EEG device and interference with surrounding electrical appliances. The most problematic group are the artifacts generated by patient, since they can not be technically avoided. Most common are eye and ear artifacts, other muscle artifacts and heart beat artifacts. The effort of removing artifacts from EEG signals is extensive which is underlined by the topic of this report. The popular approach of removing artifacts from EEG signals is achieved by using Blind Source Separation algorithms as it is more described in the section 2.2.1.

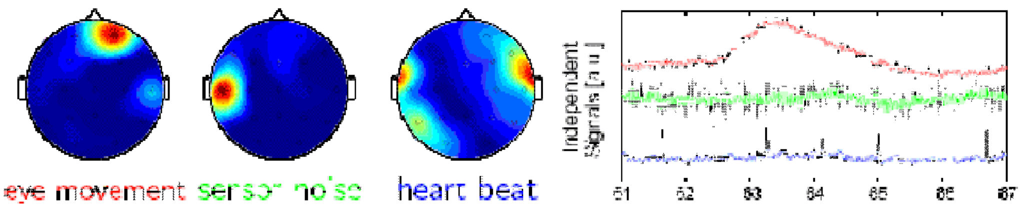

Figure 2.2 Examples of Artifacts

2.1.3 Event related potentials (ERP)

Event related potentials are a sub group of the EEG signals. They are generated as a reaction on the external or internal stimulus, clearly apparent in the signal and usually distinguishable on the EEG by eye. The stimulus can be visual or acoustic. There can be also demanded a certain reaction by a subject. ERP are categorized according to the latency of occurrence, typical shape and main peak positivity and negativity. The most significant ERP signals are P300 [10] and CNV waves [9]. Some of the peaks have systematic identifiers, which consist of letter P or N, reflecting the positivity and negativity, and a number denoting the latency of expected reaction. The ERP signals are keeping the great attention of researchers since it is used for detection of various disorders [9], brain computer interfaces as well as lie detectors.

2.2 The original SLA algorithm

The assignment of project presented in this work as well as process of its development, have arisen from master scientific assignment currently held at TEI. This original project, dealing with EEG and Magnetoencephalogram (MEG) processing, is run by Brain Special Interest Group of TEI Crete (BSIG) [2] as a part of large network of excellence BIOPATTERN [1] incorporating circa 30 European universities. The leading part and arrangements for BSIG lab are ordered by researchers at University of Sheffield. The research tasks, proposed by their research group, are divided between several universities, situated mostly in Mediterranean area, e.g. Crete and Malta. The main idea of this cooperation is to solve and improve ways of research and processing automation of EEG and MEG. BSIG labs took a place in this project by investigating the analysis of EEG source locations, their classification and rejection. The results are supposed to be used by other members of BIOPATTERN group, in medical and technological research.

In addition to this basic introduction, it is important to note, that our project deals with sub-group of EEG signals, the event-related signals, in EEG called event-related potentials (ERP) or evoked potentials (EP) [9] [10].

The more detailed, but still brief, description of original project is explained in following sub-sections. From now on, BSIG project for EEG source localization and classification will be referenced as SLA (source localization algorithm).

2.2.1 Blind source separation (BSS)

Blind source separation (BSS) aims to recover a set of unobserved signals or sources using only a set of observed mixtures. The signal is recorded as an output array of sensors, where every sensor receives a different combination of the source signals. The general formulation of BSS problems does not assume any prior information about the original signals nor about the mixing process. To solve the given problem, the compensation conditions are set: The source signals are statistically independent and they are mixed linearly.

We can define Independent Component Analysis (ICA) of a random $\mathbf{x} = [x_1, ..., x_n]^T$ as an invertible transformation $\mathbf{y} = \mathbf{Vx}$ where $\mathbf{V}$ is a matrix chosen so that the dependence between the elements of the transformed vector $\mathbf{y} = [y_1, ..., y_n]^T$ is minimized. The sources of the ICA model are expected not to be random. Random variables are assumed to have gaussian distribution. The principle of ICA is to find linear mixtures of nongaussian variables.

The ICA projection matrix $\mathbf{V}$ is the separating matrix $\mathbf{V} = \mathbf{B} = \hat{\mathbf{A}}^{-1}$ that we are looking for. In order to estimate the ICA projection matrix $\mathbf{V}$ we must define a suitable dependence (independence) measure or contrast function and minimize (maximize) it with respect to V. Most ICA contrast functions are derived using the Maximum Likehood (ML) principle. [21]

EEG measures the electric potentials on the surface of the scalp. However, these potentials are only linear mixtures of the real sources, which cannot be measured directly or separately since they occur on different places inside of humans head. By extracting and describing the hidden sources in the brain, the scientists can achieve new knowledge and new proposals concerning the brain activity, which can be used besides the field of medicine in IT area as well (e.g. in brain computer interface [3]). The research has proved that by applying BSS methods on EEG signals, it is possible to separate the sources of electric potentials in a brain, moreover to identify artifacts, troubling the signal research, and remove disturbances from the original recordings [4]. New BSS approaches improve concurrent EEG analysis methods, which are already widely used in clinical medicine.

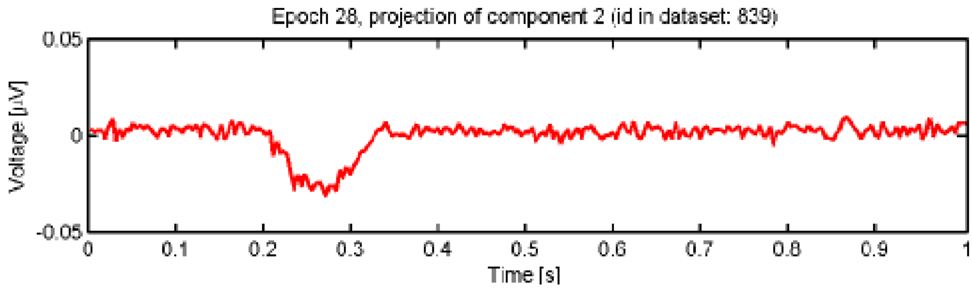

Figure 2.3 Example of ERP source aquired from EEG signal by BSS method (artificial dataset).

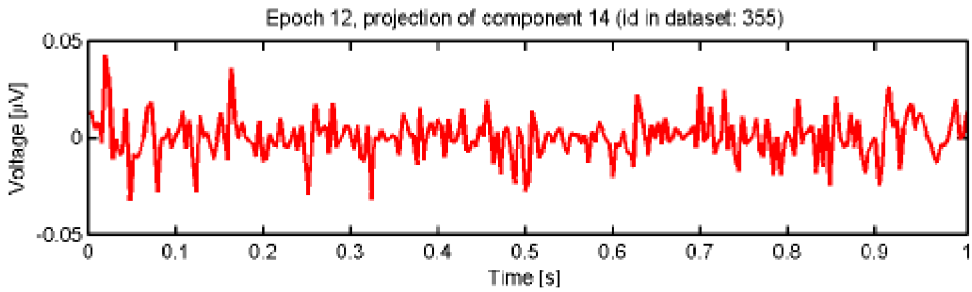

Figure 2.4 Example of artifact discovered by BSS method, which is spoiling the original data (artificial dataset).

The most common problem in the study of EEG signals is the great amount of artifact signals with large variety of sources of different origin (Figure 2.2). **Although the BSS methods are capable to separate the meaningful sources of the signal from the disturbing ones (in a particular way), the identification of real brain sources and artifact components is performed visually by specialists, rather than automatically by computers. The automatic artifact**

identification is object of current research. The demand of automatic classification is increasingly rising with the amount of recorded data and subjects.

One of the most commonly used BSS approaches is application of Independent Component Analysis (ICA) algorithms (the other is e.g. Maximum Noise Fraction [5]). Many varieties of ICA such as InfoMax [6] or Jade [7] are developed and discussed in the scientific community. The output of ICA is in the form of unmixing matrix, which expresses the linear weight of located independent source at every potential channel. As an opposite, by mixing matrix (inverse of unmixing matrix), we can project one or more component signals back through one or more channels. Following this approach, we can for example achieve cleaner EEG signal without artifacts. Since the artifacts behave as signals, that are independent from signals produced by brain activity, ICA is able to recognize and separate them. After identification of artifacts, one can project back (by using mixing matrix) non-artifact components only and achieve EEG signal in the form of channel recordings without artifact sources recognized by ICA. The component weight matrix contains spatial information as well, if we know the location of electrodes on the scalp. In our SLA approach, we are grouping the components based on their signal properties (obtained by ICA) and spatial location too. The detailed explanation of theoretical fundamental principles of ICA algorithm is beyond the scope of this report and a Figure 2.3 and adequate references [8] are provided only.

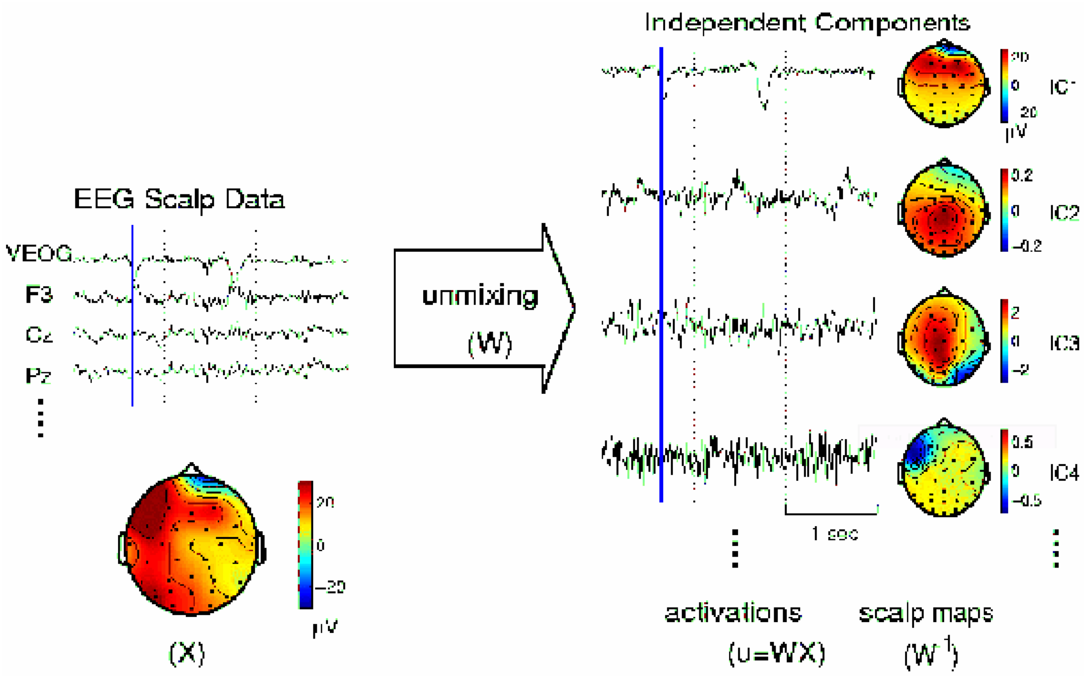

Figure 2.5 Principle of ICA localisation. The scalp maps on the right are representing different electric sources in the same time frame. Multiplying original data and unmixing matrix results to independent component signals (IC1, IC2, ...). By multiplying every independet component with a column vector of the mixing matrix (W^{-1}) one can obtain the potential of the component generated on every electrode (scalp maps on the right side of figure).

2.2.2 Scheme of Source Localization Algorithm (SLA)

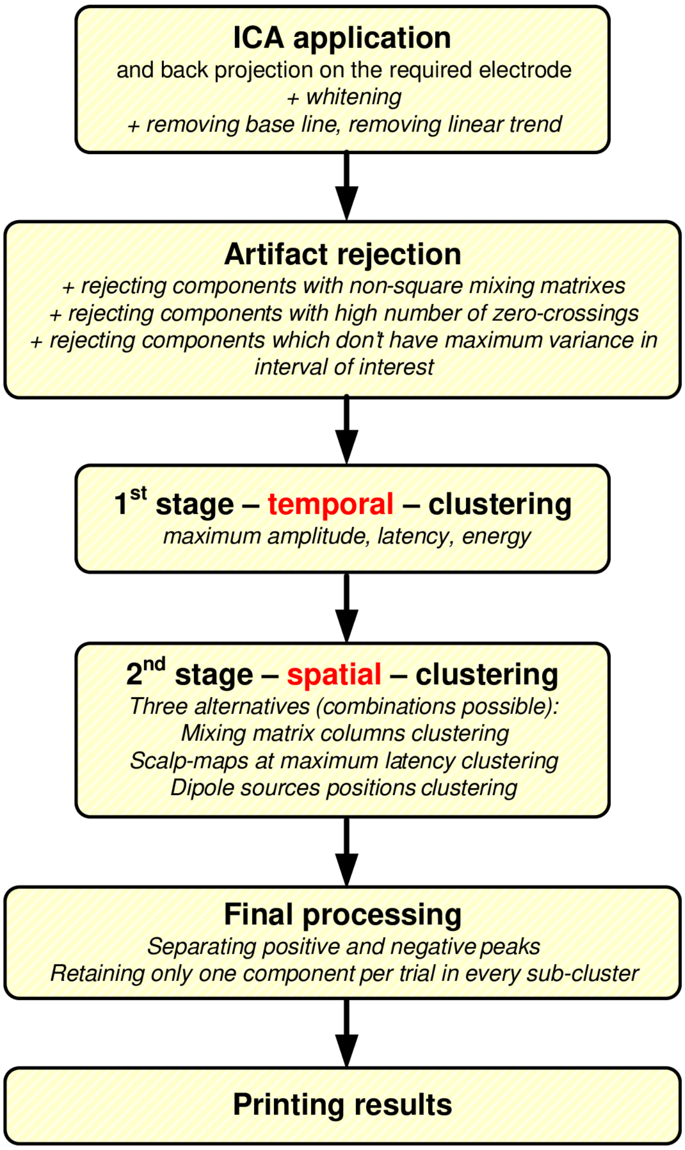

Figure 2.6 Scheme of SLA

2.2.2.1 Initial operations

All the conclusions and proposals of this report come from the SLA algorithm. Thus, it is essential to explain its basics principles due to its starting point role in the process of concluding the next steps and results.

The overall scheme of SLA is shown on the Figure 2.6. **The aim is to discover hidden sources of EEG signals and separate them to different groups based on their type and source of origin.** The artifacts should be separated from ERP signals prepared in this way for consequent medical research. ERP signals have to be grouped according to their type as well.

The first step of SLA is characterized by application of ICA algorithm. Brief description of its principle is explained in previous section. The point of this approach is to obtain and locate the sources of the electric signals before their grouping. There were few variants of algorithms tested on EEG and the most promising and powerful one seems to be InfoMax [6] algorithm, which is currently used for the BSS. The sources are not studied directly as the outputs of BSS algorithm. They are projected back on selected potential channels, in case of our group research, usually each source on one channel only. The channel location for back projection depends on the requirements and demands of researcher, who requests our data analysis. The most common projection is through CZ electrode, as the researchers believe, the most active brain area, which produces ERP signals, surrounds this electrode [10]. The system of naming and numbering electrodes [11] we use is described on the Figure 2.1.

To improve the performance of ICA algorithm by uncorrelating the components, before the computation, the data are whitened. This means that the observed vector is linearly transformed, so that a new white vector is obtained, (i.e. its components are uncorrelated and their variances equal unity) [8]. The whitening process as well as the ICA algorithm is provided by the University of Sheffield in the form of MATLAB toolbox.

Before performing clustering and classification algorithms on back projected signal sources, new preprocessing routines are applied on the data. The base line value is removed from original signal (centering of the signal in the dimension of the voltage axes to zero value) as well as linear trends. Linear trends are removed in the same manner as in the standard MATLAB function, where the best straight-line fit linear trend is removed.

In the next phase, the samples, that resulted in wrong unmixing matrix after ICA application are removed. The ICA algorithm should result in square-shape unmixing matrix only if data are correct and sources within the data are truly independent. Non-square unmixing matrix samples are not treated as clearly separable and are rejected from the sub-sequent processing.

Zero crossings and maximum variance intervals criteria are then applied to remove most easily recognizable artifacts from the dataset. The simplicity of these methods makes it impossible to remove disturbances that are more complex such an eye movement or hearth beat (Figure 2.2), nevertheless they successfully reject high amount of data, which are not decreasing the performance of following steps calculation, which is desirable. The approach of rejecting data based on their zero crossing value can result in unwanted loss of ERP components, thus it has to be discussed and substantiated in its cases of use. By a term, "maximum variance", we mean definition by Aapo Hyvärinen in his work [8], which describes it as the time frame of a signal with highest absolute value of its amplitude. At this level, up to circa 60% of the data samples are usually rejected.

2.2.2.2 Clustering phase

The first stage of clustering is applied on the set of features, which are reflecting the temporal properties of the data and their derivations. Although the valuable and meaningful data are grouped together with the artifacts, this stage has the most significant role in the final shaping of group clusters. The feature with most significant impact on the clustering is the latency, which can also cause problems in certain datasets as it is described in section 2.4. The selection of features problem was discussed many times and time frequency features investigation was proposed.

The second stage of clustering is aimed on the final removing of artifacts and classifying the data according to their location within the brain. Two most effective methods are used to solve this problem: mixing matrix columns clustering and dipole localization. Mixing matrix columns reflect the weight of ICA sources at every EEG channel and, in our case, include the information in 2D coordinate system. Dipole localization approach is trying to discover potential dipoles of electric sources within the brain, in spatial 3D coordinates. As an alternative, the experimentation with clustering of back projected scalp maps in the maximum variance time is available, but seems not to be as useful and advantageous. Besides the grouping of desired signals, the secondary role of this phase is to finally separate signals from artifacts on their spatial bases, since they should have different characteristics.

After the final data manipulations, the results are usually printed and new conclusions are drawn. The algorithm is than adapted and improved to result better outputs if necessary. The great amount of the aspects needs to be discussed before publishing the results including the amount of applicable data and the reliability and quality of the results. The final analysis have to be done in the cooperation with medical researchers. These analysis are not part of report, since these results are analyzed on other place (University of Sheffield) and the final changes and improvements are suggested from that place only.

2.2.3 SLA results

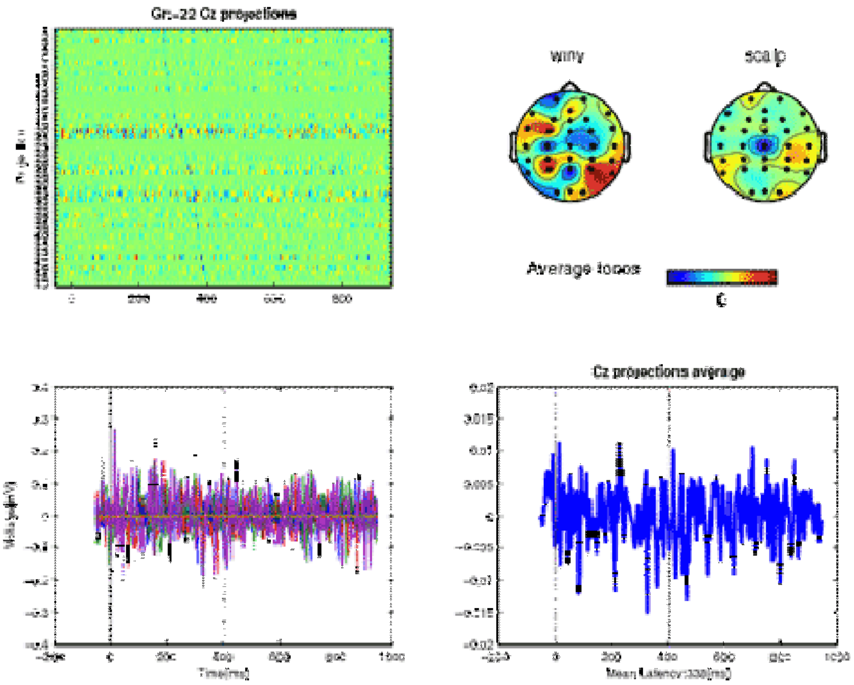

Figure 2.7 Grouped artifact sources projected on Cz channel outputed by SLA algorithm.

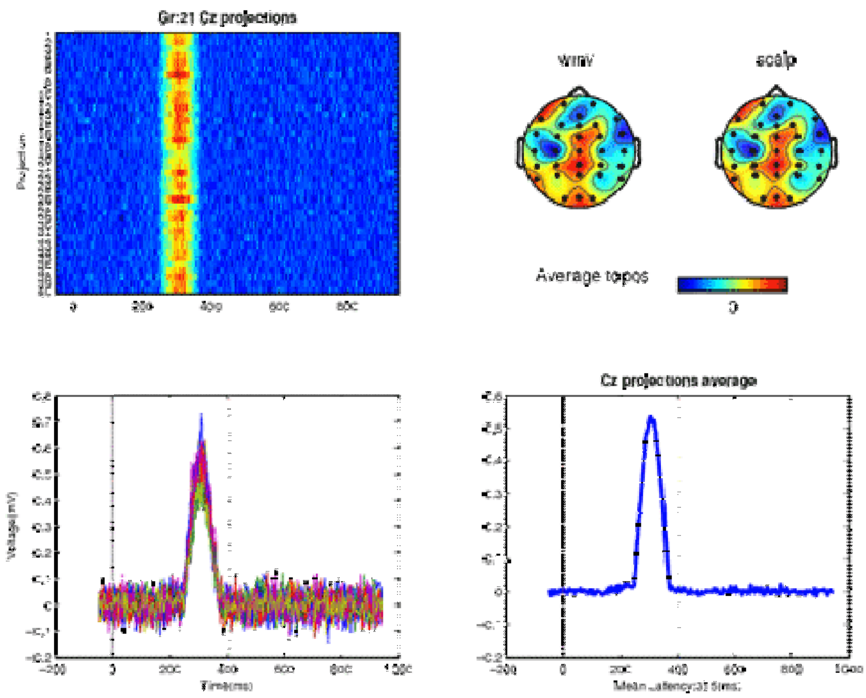

Figure 2.8 Grouped non-artifact sources projected on Cz channel outputed by SLA algorithm.

Besides the fundamental statistics and processing step analysis, the final clustering (grouping) results are plotted and visualized. The examples of typical outputs are shown on Figure 2.7 and Figure 2.8. Pictures in the lower parts of figures show examples of grouped signals and their mean signal on classical two-axis plots. Signals should have resemble the similar shape. The upper parts include all signals plotted on the same map (upper left picture), where the color determines the amplitude in every time moment. The signal map also contains a group identifier in its title, where the digit sign determines the positivity or negativity of peaks in maximum variance time within a group. The first digit informs about the first stage (temporal) clustering result and the second digit informs about second stage (spatial) result. Finally, two scalp maps visualizations show the average of mixing matrix columns of components ("winv" identifier) and their average projection in the maximum variance time ("scalp" identifier). The technical meanings of these terms are more precisely described in section 2.2.1.

The functional meaning, in other words "what are these statistics good for", is more difficult to explain and is particularly meant for medical researchers rather than for technicians. Concluding from the previous paragraph, one can say, these results are expressing the typical shape of desired signal group and its typical spatial localization within the brain.

It is important to note, that the method is in development and improved with every new dataset, it is tested on. **The aim of project, this work is part of, is to discover new features and views on these problems.** The clustering stages should be improved by discovering new and reliable features of signals. The real EEG signals are variable and unstable and the more numerical methods and characteristics their analyses have the more accurate results it can offer.

2.3 Time frequency approaches

2.3.1 Discrete Fourier Transform

The Discrete Fourier Transform is the basic approach of interpreting the frequency representation of the given signal. It has reached innumerable application in the mathematics, physics and natural sciences. The extremely efficient Fast Fourier Transform algorithm makes it very attractive computational tool and a basic pillar of the signal processing approaches.

The aim of the Fourier Transform is to express the given continuous signal as a set of harmonically related complex exponentials with different phases and amplitudes. Convoluted complex conjugate exponentials are the cosine harmonic waves with identical amplitudes and represent the original signal in the more comprehensible way. The Fourier series can be expressed in the form

$$x(t) = \sum_{k=-\infty}^{+\infty} c_k e^{jk\omega_0 t}$$

The term for k = 0 is a constant. The terms for k = 1 and k = -1 both have fundamental frequency of ω_0 and are collectively referred as the fundamental components or first harmonic components. The two terms for k = 2 and k = -2 are periodic with half period of fundamental components and are referred to as the second harmonic components. More generally, the components for k = +N and k = -N are referred to as the Nth harmonic components and have N times higher frequency than a fundamental component [19]. The $\omega_0=2\pi/T$ is the fundamental frequency as the T is the fundamental period. The fundamental component has the lowest frequency (highest period) in the frequency representation spectra of the given signal.

The Fourier Transform requires the mathematical representation of the input signal, which makes it impossible to use in our case, since our signals are in the form of sets of values. For our purposes, we will use the Discrete Fourier Transform, which is performed for analysis of discrete signals.

2.3.2 Short Time Fourier transform

Since the Fourier Transform represents the given signal spectra in the whole time domain, it lacks the information about the time evolution of frequency exponents. Moreover, if the significant signal peak occurs in the signal, the whole Fourier spectrum is affected, which can be unwanted effect. In the cases of EEG, the signals are characterized by frequent time varying features, which attract the main interest of the scientists. Thus, the time frequency resolution is required.

This problem can be particularly solved by the Short Time Fourier transform concept. The signal is separated to certain segments and the Fourier Transform is than calculated for every segment separately. The time resolution is high, which makes the method attractive for time frequency analysis. Its nature of non-limit boundaries can make it more beneficiary than a wavelet transform, which is faster, but divides given signal in fixed blocks. Another limitation is the frequency range of the representation, which depends on the segment size. Thus, the right ratio needs to be found between the time and frequency resolution. This can be a challenging task in the analysis of ERP signals, since the amplitudes of the ERP waves have low frequencies.

The time frequency analysis is one of the basic tools for detection and analysis of epileptic behavior.

The basic pillar of our approach will be the Fourier Transform (FT) algorithm. As it was already described, the algorithm is trying to describe given signal by set of cosine waves of different frequencies, amplitudes and phases. The output of algorithm is in the form of array of complex values, called coefficients. Every coefficient corresponds with certain frequency and includes the information about the phase and amplitude. As more samples we use for FT algorithm as more coefficients we get. FT algorithm demonstrates the best performance on periodic and periodic like signals. However, it can be a powerful tool for analysis of non-periodic stochastic signals, as it is in our case, too. Its application in this way needs careful consideration since the high amount of samples can result in highly distorted outcome. The FT coefficients do not include any time information. This lack can be overcame by applying the FT algorithm on segments of the signal, which can result in better resolution than another widely used time frequency analysis approaches e.g. wavelets. Finally, in our solution, we will use a FFT algorithm, the efficient algorithm with favorable complexity.

2.3.3 FIR filter

A filter is essentially a system or a network that selectively changes the input signal. Filters are commonly used for improving the quality of the signal, extracting the information from the input signal or separating two signals from each other. Finite impulse response (FIR) filter is characterized by the finite duration of its response. Filter is characterized by the set of parameters that express the filter specification. Most of the filter specifications refer to the frequency response of the filter. The filter can be than able to pass only desired frequencies. This technique is used in our approach for detection of artifact of high frequencies, which are not outputted by used filter system.

The usage of FIR filters does not play a crucial role in our approach thus the more detailed explanation will not be given within the content of this work and an adequate resource [18] will be provided only.

2.4 Time frequency analysis for SLA

Despite the facts, that frequency analysis is the most common approach of investigation of EEG, it has neither been integrated into the original SLA paradigm and nor discussed within its context. With a time, when we were looking for improvements and solutions for SLA, which would conclude in more stable system, the necessity of study of frequency features of EEG had arisen. Along with the looking for solutions of already known problems (Figure 2.9), it was necessary to investigate the time frequency analysis and explain its usefulness in the context of the SLA algorithm.

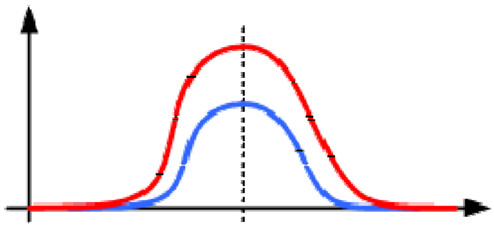

Figure 2.9 The problem, which can not be resolved by the orginal algorithm.

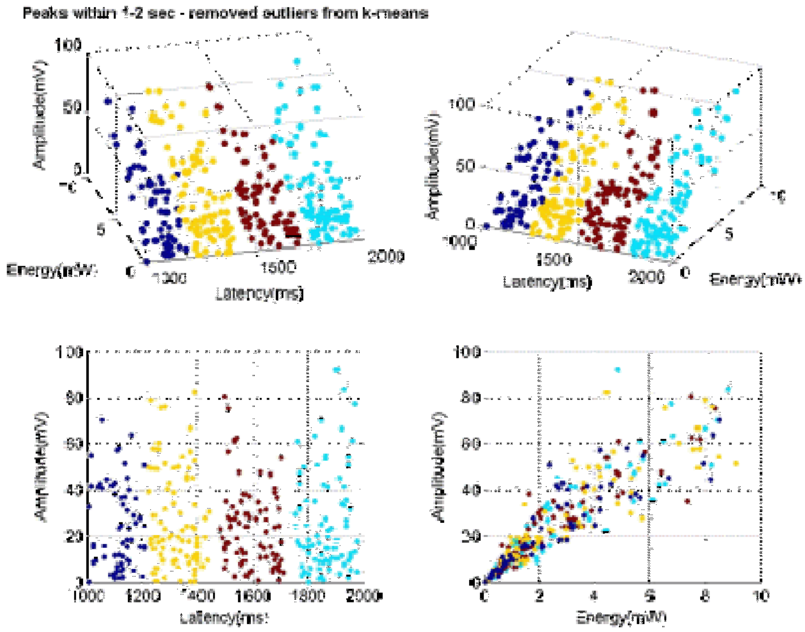

Figure 2.10 Spatial relation between clustered features, the first stage of clustering during SLA development.

The Figure 2.9 shows a possible problem of the SLA approach. Two signals could be grouped to the same cluster, since they have the same latency in maximum variance frame. The deviations of energy and amplitude are not high enough to place these signals into different groups. Frequency analysis method seems to be a good way for separating these two peaks. The problem of dominance of latency feature (Figure 2.10) shows a necessity of exploring new features, which can be more advantageous to use for signal clustering.

3 The proposed solution

The proposal of our approach comes from the problems and challenges, which are meeting the SLA algorithm and which are discussed in previous sections. The proposal is in the form of an algorithm, which defines the possible use and way of application of discussed methods. Because, the SLA algorithm is permanently under the development it was not possible to demonstrate the methods as already integrated newly designed parts of the master implementation. The time frequency analysis study should serve rather as a reference material than a solution or a part of original SLA. On the other hand, the usefulness and usability of new methods is necessary to be agreed by other scientists and as an integrated part of SLA it could have been a better way of promotion.

As a separated entity, the flow of ideas and conclusions will be independent, and interesting topics and proposals for SLA improvement will be suggested and commented during the explanation of implementation. In the end, the final summary of these suggestions and proposal of future work will be given in the section 3.7.

The approach using a Fourier Transform was chosen, because it is the most commonly used method for frequency analysis in general as well as in the case of EEG signals. **We will try to build the algorithm that recognizes different types of ERP signals based on their time frequency analysis only.**

3.1 Data and implementation framework

The development of project was initiated by the work on artificial data provided by University of Sheffield. After the concept was finished and implemented, the testing on real EEG dataset was following. The details of datasets and results are published in the appendix part of this report.

The implementation of proposed ideas is programmed in the MATLAB 7 framework due to the compatibility with original SLA algorithm and its suitability for signal processing solutions.

The **artificial** ERP dataset was obtained from the University of Sheffield in the form of MATLAB matrix file. It contained 40 recordings of 31 channels EEG (1240 signals). Samples 1s long were acquired with sampling rate of 250 Hz. The interval of interest, where the ERP peaks were searched was 50-456 ms. After first stage artifact rejection, 1080 samples were rejected and 160 kept for subsequent computing. STFT was applied in 90 windows. Every window had size of 40 frames. Data were grouped to four groups, according to the latency of their most significant peak. First group (with index 0) contains peaks on the borders, which were found to be out of interval of interest. Signals were clustered according to their first four FFT coefficients into four clusters.

The **real** ERP dataset was obtained from the University of Crete in the form of MATLAB matrix file. It contained 40 recordings of 27 channels EEG (1080 signals). Samples 1,0547s long were acquired with sampling rate of 1024 Hz. The interval of interest, where the ERP peaks were searched was 568-1036 ms. After first stage artifact rejection, 577 samples were rejected and 503 kept for subsequent computing. STFT was applied in 350 windows. Every window had size of 160 frames. Data were grouped to four groups, according to the latency of their most significant peak. First group (with index 0) contains peaks on the borders, which were found to be out of interval of interest. Signals were clustered according to their first 10 FFT coefficients into four clusters. Some of the result plots are present in the Appendix B.

3.2 Concluding the algorithm of time frequency analysis

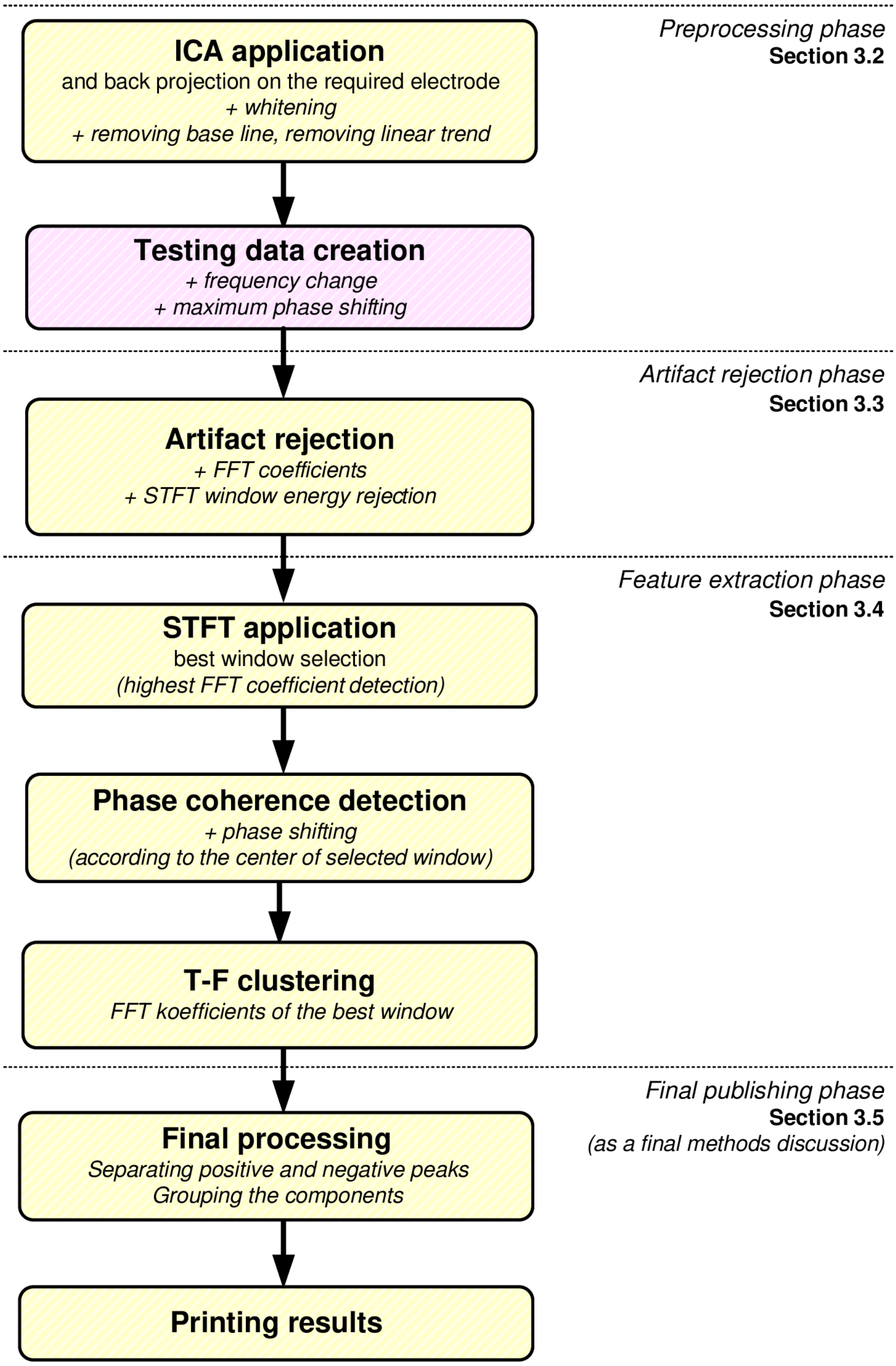

Figure 3.1 Proposed sequence of steps for demonstration algorithm of frequency analysis of ERP signals

Frequency analysis and time frequency analysis are quantitative methods of describing oscillations [17] in general, the electric signals in our case. This attribute is marking out its possible role in classification of ERP signals. The main attention will be paid to use it as a tool for rejecting artifact signals from EEG samples and mining the features for temporary and quantitative grouping rather than spatial information search. The study of frequency features and their explanation will be given based on the algorithm. The scheme of the algorithm is on the Figure 3.1. The conclusions are drawn on the base of the algorithm results. Its implementation pieces can be used as the basic grounds for implementing the proposed features to the final solution.

The first step, in the same way as in SLA, is to locate sources of EEG signals and project them back on required electrode. This operation is essential in the SLA algorithm and was a basic requirement for this work, since we are interested in studying signals sources (their back projection) rather than original EEG signals. The analysis has to be done using data preprocessed in this way. Removing the base line of signals (centering in voltage axes direction) as well as removing the trend are common operations before application frequency analysis on any signal.

More interesting is the following step, creation of testing data. It was found, that artificial testing data at our disposal are not ideal for performing tests based on frequency analysis, because of their similar frequency characteristics. Thus, a new method was necessary to be proposed for acquiring new data with various frequency characteristics. Changing the properties of already existing projected data was set up as a most effective way since the process of generating brand new data with spatial information for input of ICA algorithms seemed to be too complicated. This approach would require adequate medical background and the work alone would be good proposal for master thesis topic. The principle of new data creation algorithm is explained in section 3.3.

Next intent is to use frequency analysis for rejecting the artifact components. ERP signals are characterized by sudden high peaks of low frequencies as some of the artifacts dominate in high frequency ranges. This problematic is examined in section 3.4, where a Fast Fourier transform (FFT) is applied on the signals and discussed. However, the real data are not characterized by sinus shape like peaks, which causes FFT inefficient on real data sets. Better approach of rejection using FIR filter is introduced in the same section.

Finally, the most interesting results are present in section 3.5, which describes an application of Short Time Fourier transform (STFT) on the dataset and its impacts on the signal analysis. The method incorporates two steps for quantification of signals: Highest peak localizations and its frequency clustering and grouping.

The final phase is designed to prepare results for presentation and perform a plotting. Since it is a technical matter of implementation and results are discussed within whole the section, it is not principally described. The overall frequency analysis contribution to SLA is rater discussed at the end of this section in the section 3.7.

3.3 Preprocessing phase and modification of testing data

Some necessary steps have to be applied before using our methods for signal analysis, especially FFT. The preprocessing methods include whitening, removing the base line and removing linear trend from EEG signals. The usage of whitening preprocessing improves ICA performance and is chosen based on the recommendations of our partners at University of Sheffield. Removing base line is fundamental step, because in the certain moments, the normalization of signals has to be done. If the signals would not be centered it would lead to the loose of information and distortion of results. Removing linear trends is commonly applied routine before applying FFT to achieve better results. Although our testing data do not include this feature, it is very common in real EEG data. All the preprocessing routines are performed in the same manner as in the original SLA approach.

With the analysis of given dataset, the demand of new testing data has arisen, since, the data do not consist of sufficient variety of samples, with different peak frequencies and latencies. On contrary, in the EEG investigation, it is the basic demand, to separate peaks with different frequencies since it is often object of research. Thus, the final solution incorporates the algorithm, for creating peaks with various frequencies and latencies.

The principle of signal modification algorithm is the following: The signals are cropped and divided into several parts with random portions of size. Then the sampling frequency of these parts is randomly changed, but time portion of segments in the whole signals scope is preserved. The sampling frequency of cropped edges is then adjusted to reach the same amount of time frames as the original signal had. The change of sampling rate in all segments results in the latency shifting of peaks. Finally, because of change of sampling frequency of the segment with the most significant peak and preserving its time period changes its final frequency also the significant peak. The effect of algorithm can be seen in Figure 3.2.

The context of signal modification algorithm and the SLA does not need to be discussed due to its testing intention. The preprocessing performed in the master solution corresponds with the presented one and the signal modification is not necessary to be done since it serves as a way for showing the method performance.

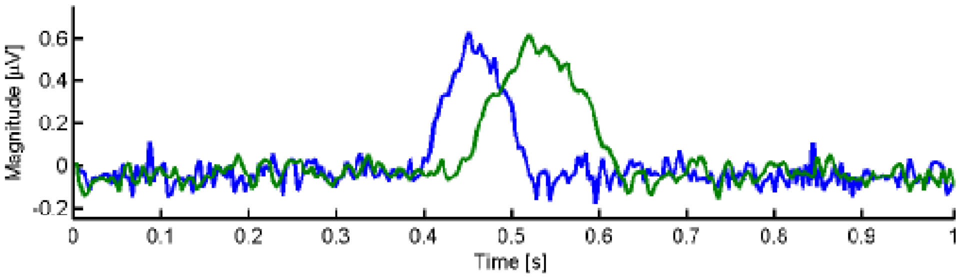

Figure 3.2 Detailed example of original signal (blue) and new signal (green) acquired by the change of sampling rates.

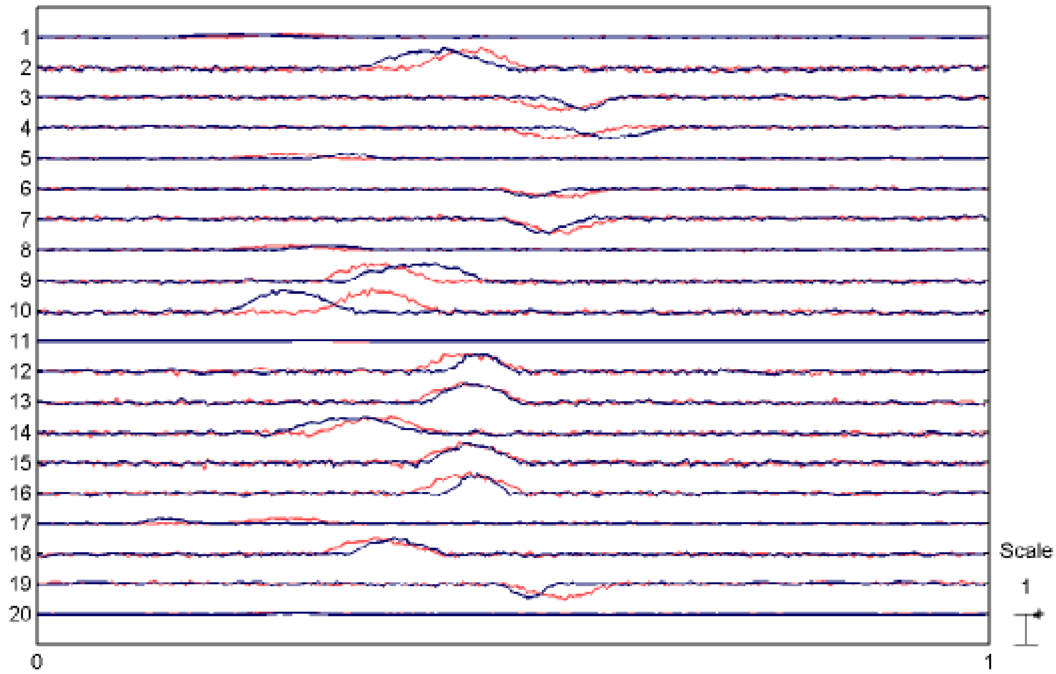

Figure 3.3 Examples of new signals (red) acquired by sampling rate changes

3.4 Rejection of Artifacts

Probably the most discussed topic in the theory of processing and interpretation of EEG signals deals with the artifacts, their identification and avoiding. As it was described before (and shown on Figure 2.2), many types of artifacts spoiling the EEG signals may occur. Although various methods are proposed for automatic removing, the final decision has always to be done by a doctor or researcher. Since we are using clustering methods, the stage of early artifact removal from dataset can have great impact on performance of grouping algorithms in the later stages of processing. That makes the artifact rejection an important step, which has to be carefully considered especially in the context of our approach, where it was found crucial for successful frequency features clustering.

3.4.1 FFT approach

In our artifact rejection approach, we will try to focus on removing all not-like-ERP artifacts by using whole signal FFT spectra information. ERP artifacts are characterized by significant positive or negative peak observable by eyes.

All the signals, which do not include this kind of peak in our interval of interest (interval, where we are expecting ERP reaction) are rejected first.

One can easily distinguish on Figure 3.4 and Figure 3.5, how different the frequency characteristics of displayed artificial signals are. The FFT coefficients have different characteristics and are easily separable.

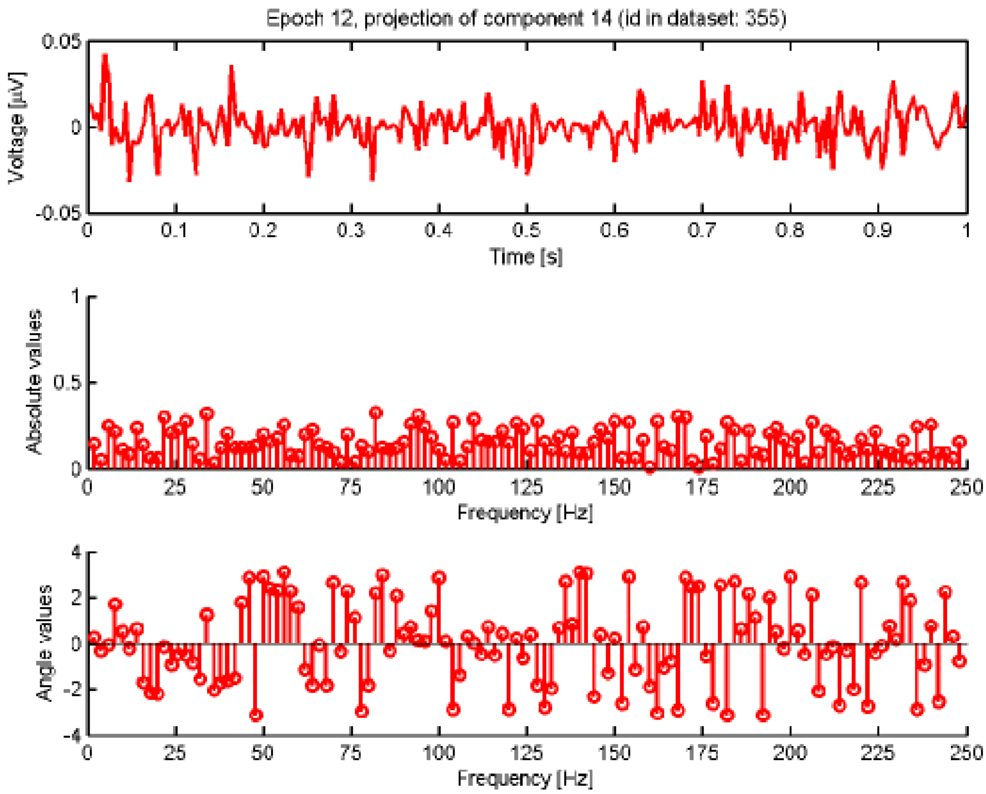

Figure 3.4 Fourier Transform coefficients of artifact signal, adjusted view

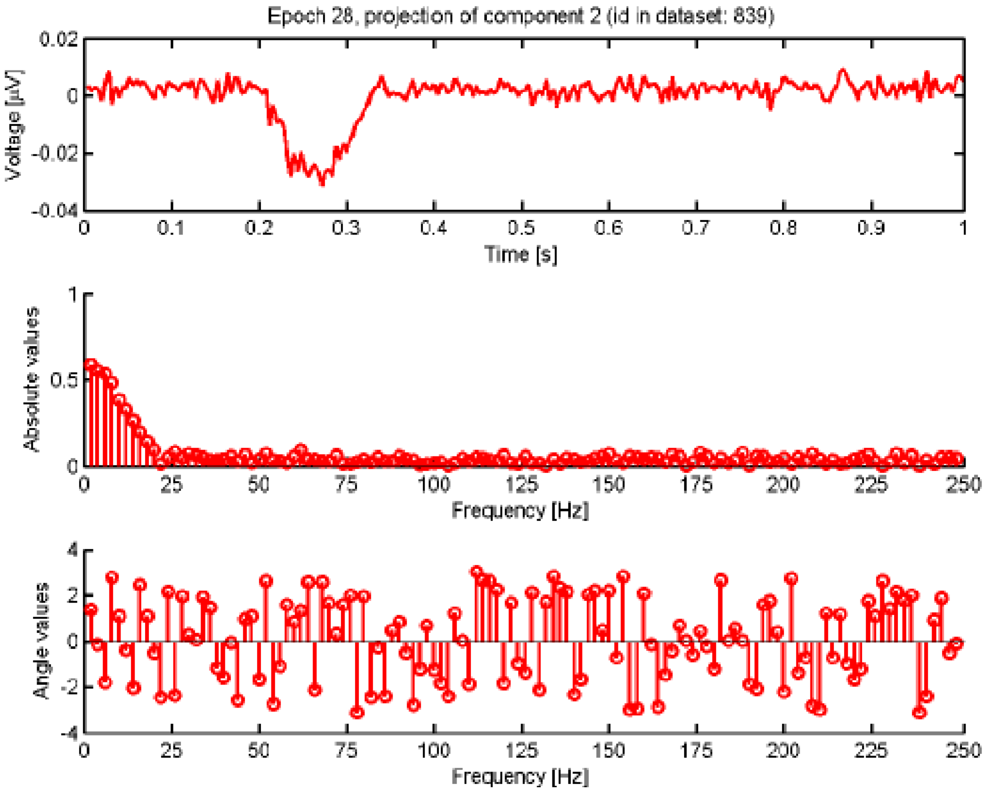

Figure 3.5 Fourier tranform coefficients of ERP signal, adjusted view

Another problem concerning the values of absolute FFT coefficients has arisen during their grouping. Different amplitude of their peaks can cause significantly different coefficients and influence component grouping. The problem has been solved by normalizing coefficients in whole spectrum. For more details, see Figure 3.6, Figure 3.7 and Figure 3.8.

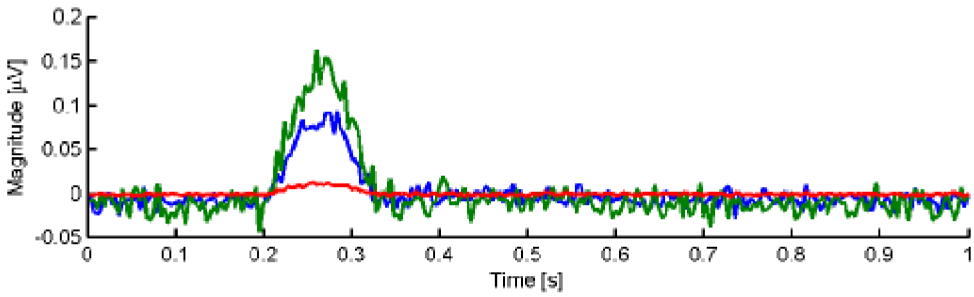

Figure 3.6 Three components, with the highest peak of similar frequency but different amplitudes.

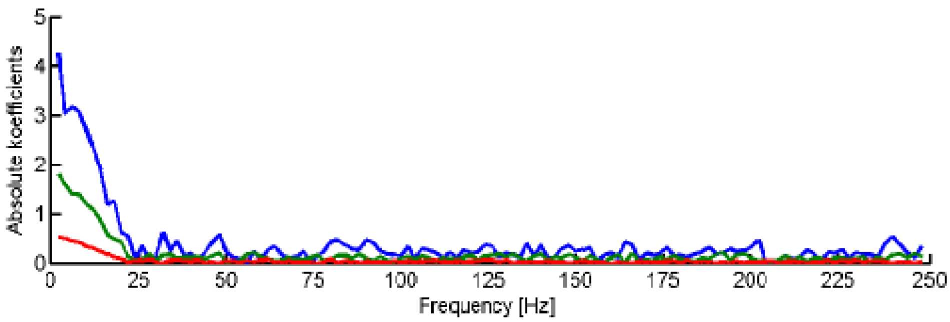

Figure 3.7 Absolute values of FFT coefficients of three components on Figure 3.6.

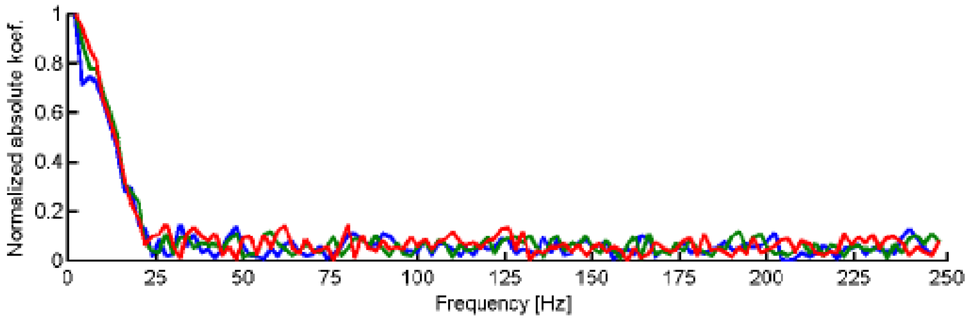

Figure 3.8 The normalized absolute FFT coefficients of three components on Figure 3.6.

3.4.2 Evaluation of FFT approach

The clustering using averaged FFT coefficients groups as features was performed and the results can be seen on Figure 3.9. The features had in case of artificial data good enough quality to be used as clustering features. The clustering was 100% successful, all artifact were removed. The classification was done on a base of five features, representing typical frequency bands in classification of EEG signals (frequency bands are described in section 2.1.1). The upper limit of gamma band in our algorithm was sampling frequency.

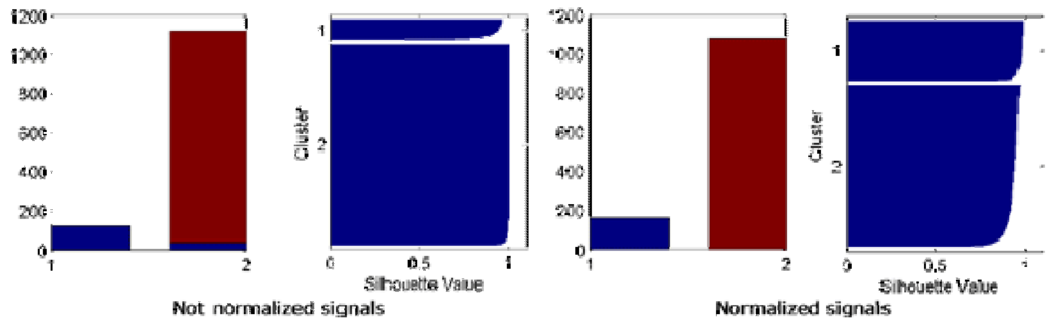

Figure 3.9 The results of clustering FFT coefficients. Blue color on box plots denotes the ERP components, red color denotes artifacts. Boxes resamble the composition of both clusters. The silhouette plots are added to show the performance of clustering. The silhouette value for each point is a measure of how similar that point is to points in its own cluster compared to points in other clusters, and ranges from -1 to +1 [15].

However, while testing the method on real data, new problems have appeared. Probably the fact, that ERP peaks do not have shape like sinus waves, makes FFT features unsuitable for successful clustering (Figure 3.10 and Figure 3.11). Neither over sampling (adding zeros in the end of signal vector before FFT application) neither down sampling (decreasing high sampling rate of real datasets) brought significant improvement to the method thus another solutions had to be proposed.

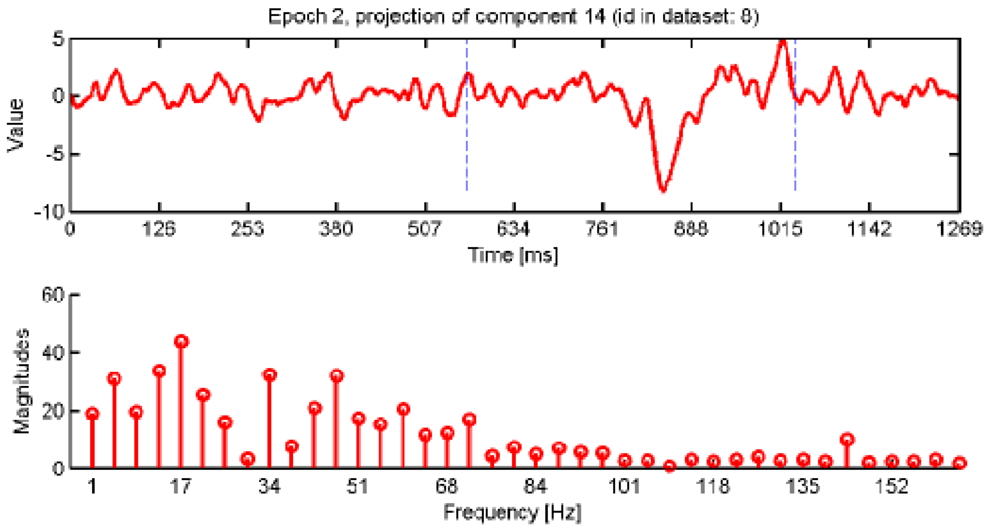

Figure 3.10 FFT absolute coefficients of the ERP signal from real dataset. FFT was applied on the signal within blue vertical line borders only. The significant negative peak starting at time about 700 ms is not reflected in FFT.

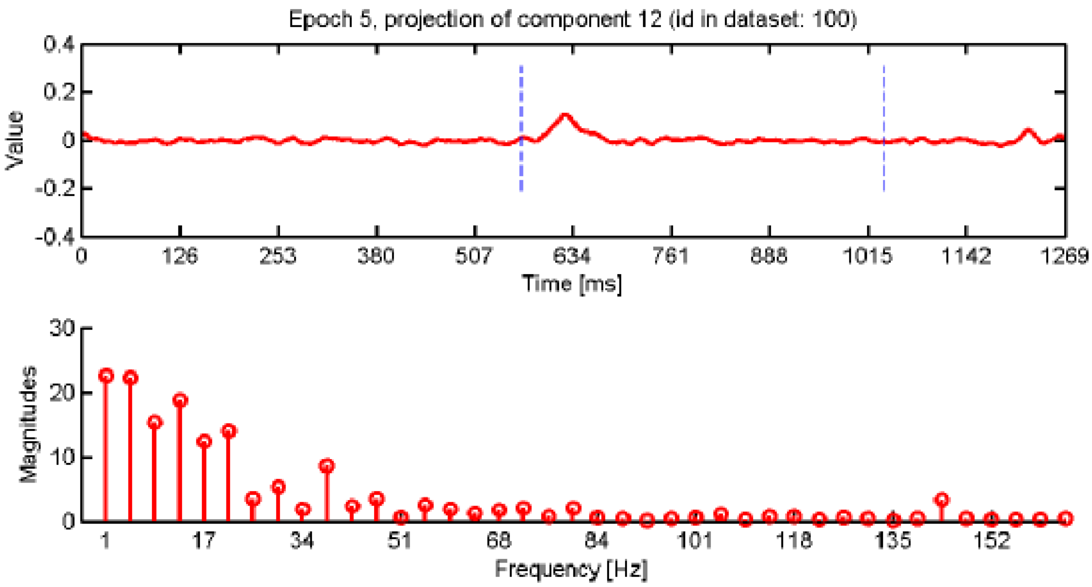

Figure 3.11 FFT absolute coefficients of artifact signal from real dataset. FFT was applied on the signal within blue vertical line borders only. Enormously high coefficients are present in low frequency ranges.

3.4.3 FIR filter approach

Finally, FIR pre-filtering was applied and results that are more promising were obtained. Signals were filtered using FIR filter, which was passing low frequencies only. Artifacts were recognized based on the standard deviation of the filtered signal. The classification by threshold was used to separate signals. The period of interest used for calculation of standard deviation was shifted because of filter response delay. All the signals were normalized.

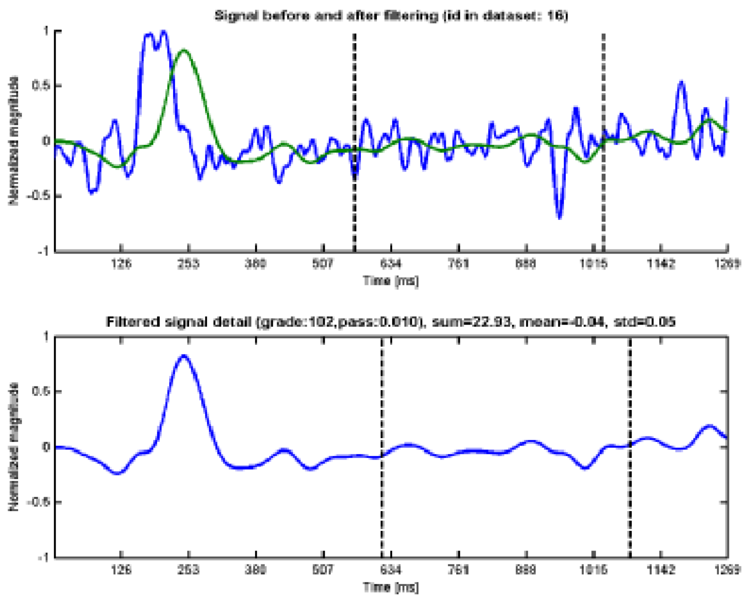

Figure 3.12 The signal rejected based on the FIR filtering method. The vertical lines define the interval of interest. Blue color in the upper part denotes original signal as the green color shows its filtered variation. Lower plot shows the filtered version only with its shifted interval of interest used for computation of statistical methods, which are present in its title.

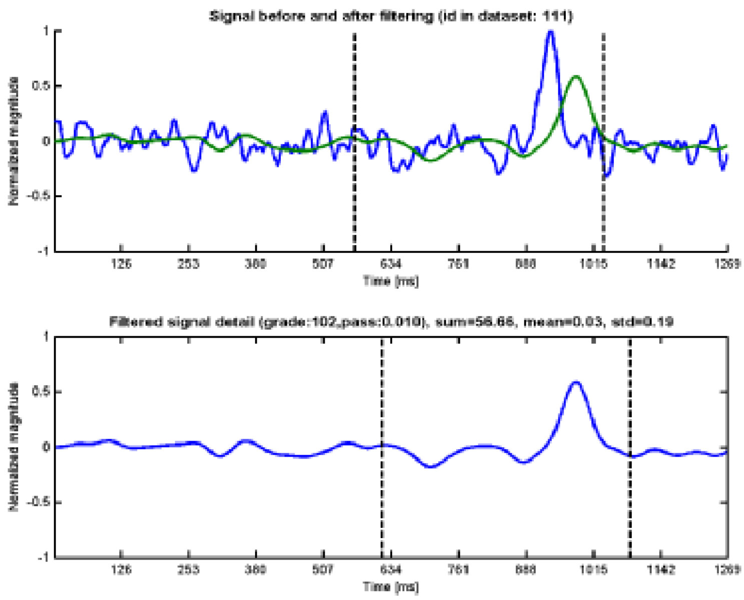

Figure 3.13 The ERP signal identified by FIR filtering method. The vertical lines define the interval of interest. Blue color in the upper part denotes original signal as the green color shows its filtered variation. Lower plot shows the filtered version only with its shifted interval of interest used for computation of statistical methods, which are present in its title.

As a conclusion, we have shown that FFT analysis cannot be used at the current stage of process. The artificial data and the shape of their peaks made FFT features inefficient. As a result, another solution was proposed based on the FIR filter. The output quality makes this method possible to be used in the SLA algorithm on the artifact rejection layer, where maximum variance rejection is currently performed.

3.5 Application of the Short Time Fourier Transform (STFT)

The main idea of proposal of this project is to present the alternative to the first step of SLA clustering, where the signals are clustered on their temporal (quantitative) base. The requirement of temporal condition requires using different technique than FFT.

In this step, the peak with maximum variance is selected by the algorithm and its amplitude, latency and energy features are used for clustering. As it will be shown in this section, application of STFT can substitute this phase of processing and makes it possible to acquire new clustering features of good quality. On the other hand it is important to note the complexity, which is significantly higher and decreases the performance greatly. However the linear-logarithmic complexity (N*log N, N-number of points) of FFT algorithm does not cause a huge work load for current computers with high computing power in context of our dataset sizes (thousands of samples). The results have also shown that the method is very sensitive to the quality of input samples, which makes it more reasonable to use the approach as a complement of the SLA algorithm rather than a replacement of certain part.

3.5.1 Search of the most significant peak

New approach of localization of the highest peak using time frequency analysis is the following: Before any spectral computation, we have found useful to normalize signals. That increases the performance of algorithm when it compares FFT coefficients of signals with each other. The FFT is applied in windows on original signal. Every window is multiplied by Hanning window before application in order to avoid border problems. Window positions are uniformly distributed within the signal vector. Their size as well as the amount of applications depends on the sampling rate, overall length of the sample vector and frequency of peaks to be localized. The computation window has to be able to capture whole the peak in order to analyze it successfully. For example, if searched ERP peaks are of frequency 5 Hz and the sampling rate of input signal is 250 Hz, the size of the window has to be at least 50 frames long in order to capture the wave of required size. From the size of the window and the length of the area of interest, where we try to search for EP's, we can conclude the amount of applications of the window, their distribution along the signal vector. The more application of windows we perform, the more accurate results can we achieve and the more time it takes. On the contrary, the overall sum of number of window applications and window size must not exceed the length of area of search, since we are working with discrete signals and we are unable to achieve better resolution. Higher frequency resolution can be achieved by over sampling the window before FFT computation, stretching it by adding signal of zero values. The more samples we artificially add, the higher resolution of frequency bands we can get. Although this strategy can bring the positive contribution, we have not found it useful in the context of our solution.

The next challenge is to find the window with the ERP wave, which is usually the most significant peak of the signal sample. That is done by searching the output matrix for the highest absolute FFT coefficient. The window, which is including this coefficient, is than selected and used for further analysis as the window containing the ERP peak. Now we will try to group similar ERP signals together according to the frequency spectra of the selected window and to its position within the sample.

All the other information used for grouping the signal and extracting features for clustering are concluded from this selected window only. The process of grouping the components may be divided into three steps: the frequency feature clustering, latency classification and peak negativity determination.

3.5.2 The search of ERP peaks

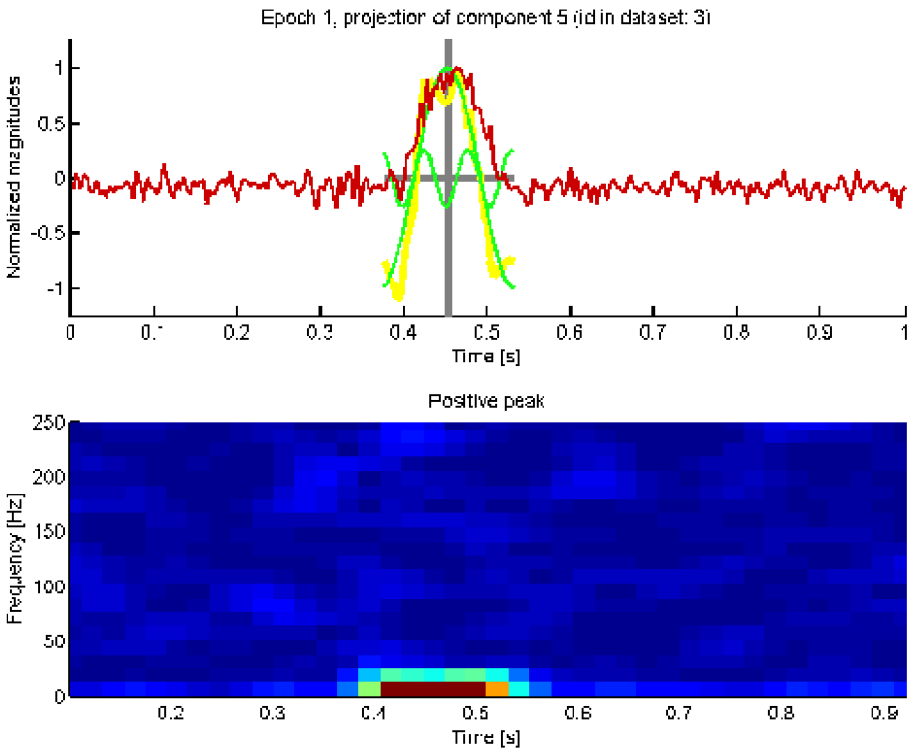

Figure 3.14 Example of the result of STFT algorithm application on the signal. The signal in upper part is highlighted by red color, while the window with most significant peak is gray. Green color denotes visualization of firts two FFT coefficients of selected window and yellow color shows the convoluted signal from FFT coefficient reconstructions. The image in lower part shows the values of all FFT window applications. Every column corresponds with a window, blue colors demonstrate lower coefficients as an opposite to high red ones.

Absolute coefficients from low frequency spectra of the selected window were used as features for the clustering. The amount of the coefficients can vary, depending on the resolution of the window and sampling frequency. Two different ERP signals are supposed to be separated since they are not expected to have similar low frequency characteristics of low frequency domination samples in their window. This clustering brings most significant contribution to created groups. The problems can arise, when sufficient amount of low frequency domain artifacts is not removed from the dataset, before clustering. In this case, the method separates the artifacts from ERP signals and it looses its sensitivity to recognize different ERP signal groups. However, this defect can be sorted out in artifact rejection phase by sophisticated algorithm.

3.5.3 Signal shifting, phase coherence

In subsequent experiment, we try to group and shift the data according to the position of their selected FFT application window, based on the peak phase coherence. Concerning the fact, that signals are clustered according to similar frequency characteristics of their selected windows, we can center all the signals in the way that their peak centers appear on the same position. This can make their shapes easier to compare visually. The time axis was divided into several periods and signals were classified depending in which period their most significant window was found. All signals of a group where than shifted according to their selected window center latency. The problem has arisen with the signals, which peak was on the border of interval of interest, and the algorithm was working with cut peak. Thus, the analyzed interval was extended to cover the peaks exceeding the area.

3.5.4 Negativity of ERP peak

The last feature of selected peak is calculated from the window, its negativity. This is done in very simple way according to the size of sum of the signal samples in the window.

3.5.5 The results

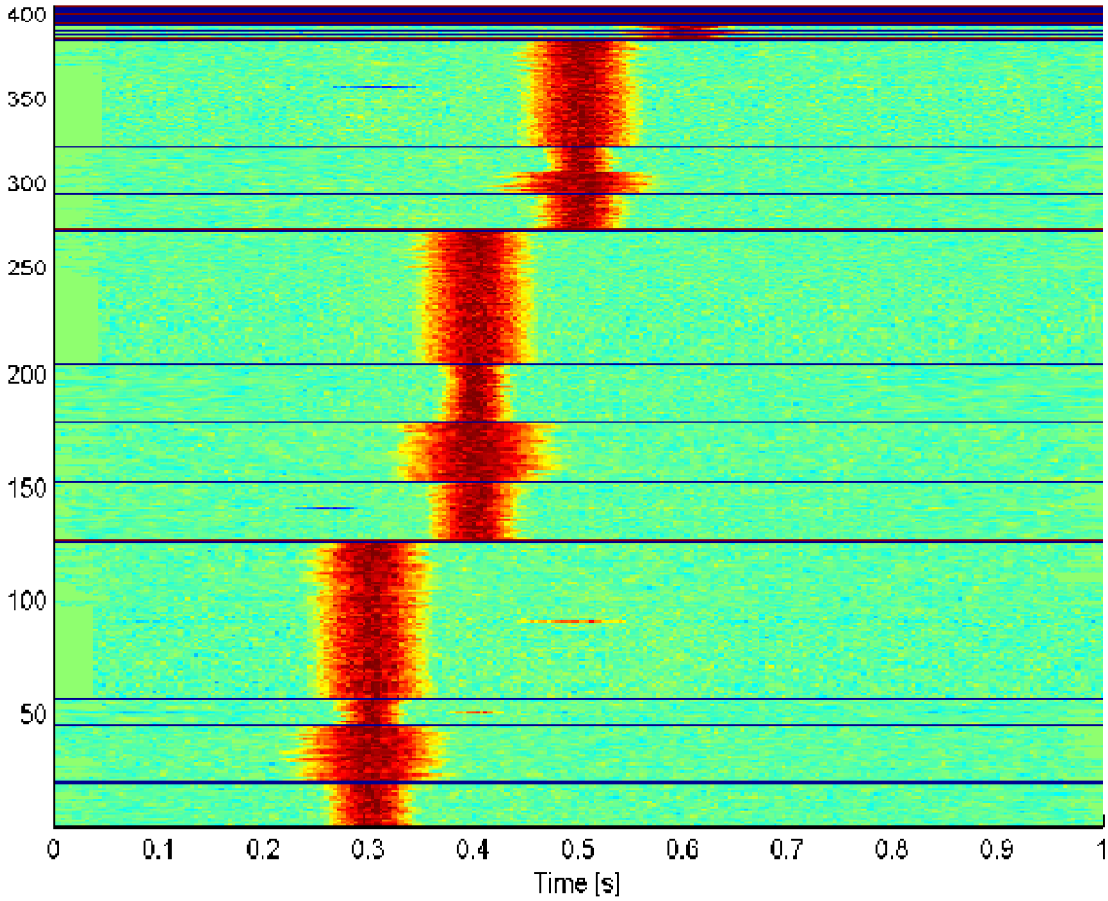

Figure 3.15 The result of grouping components with positive peak using STFT analysis. Rows corespond with signals, columns with time frames. Clusters resulted from FFT feature clustering are separated by blue horizontal lines. The sub-groups of intervals, where the peak was explored are divided by red horizonta l lines. Signals were moved so that the centers of their select FFT application windows are in the centers of latency periods.

The results of applied algorithm on artificial data can be seen on Figure 3.15. We can clearly see that the algorithm has classified the ERP peaks into four basic groups, according to the peak phase coherence. Every group contains four subgroups of signals with different frequency characteristics of the ERP peaks. This result can be described as the exemplary outcome, since the artifacts were successfully rejected and ERP signals grouped as it was requested. If the artifacts were present in the dataset (they would not be successfully rejected), the result would look more different. Possibly the artifacts would be present in one or two clusters as the ERP components in the rest. The components would not be then separated to the latency groups so clearly, since there would be more components with more different frequency ranges in the same clusters. The comparison of results can be seen on Figure 3.16 and Figure 3.18. This lack would also conclude in wrong shifting of clustered signals within their latency groups. The signal shifting is performed with assumption, that the clustered signals selected windows have similar frequency characteristics.

In the Figure 3.16 are the results of computation of real dataset. The algorithm was successful in this case as well and successfully separated and sorted the majority of ERP components. However, the small errors in phase coherence detection can be present, since the algorithm works with the real data. The errors have occurred in the latency level of 500ms, the upper group and 800ms level, second upper group.

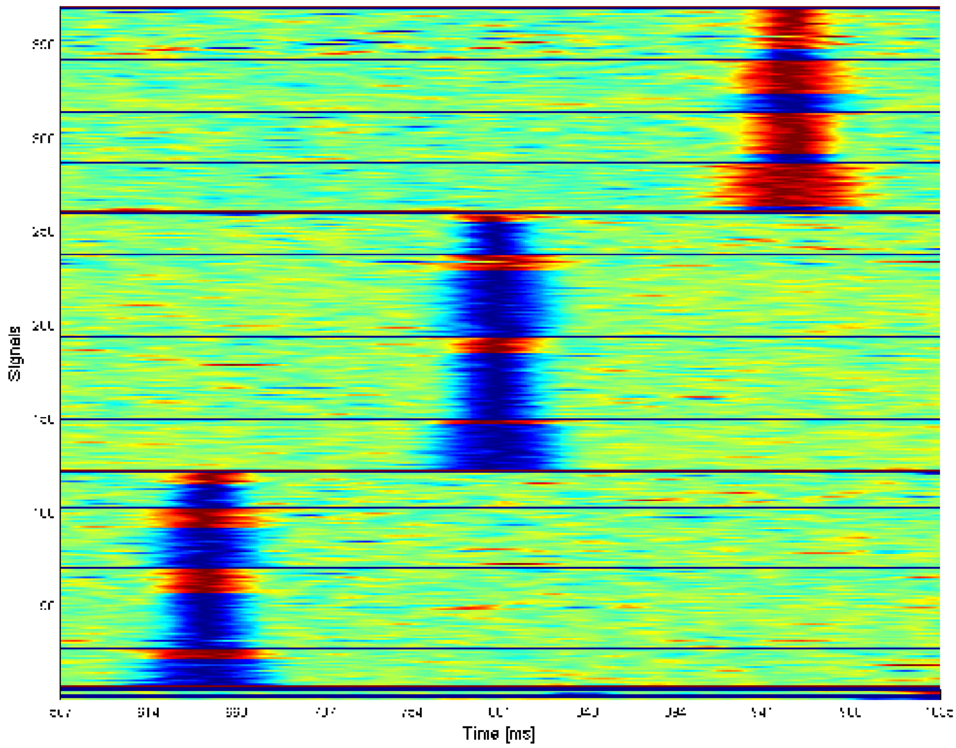

Figure 3.16 The results of the method performed on real dataset.

Finally, the method has shown that it can be a powerful tool for grouping the signals. It has proven that frequency features and methods for their mining have good potential for ERP separation. They can be successfully used for our goals. The STFT analysis can supplement the first stage of clustering in SLA to group the ERP signals and possibly separate them from complex artifacts (e.g. eye movement) without spatial information. Moreover, it can offer new functionalities, which can increase the performance and capability of current algorithm during the application on real data. Although it shown a good performance on real data too (presented in appendix section), the method has not been tested on the high amount of real datasets and has not been commented from the medical point of view. The only results for comparison we have are the ones gained by SLA algorithm. In their context, our new approach results in positive manner, since the ERP's seem to be detected and grouped in a good way.

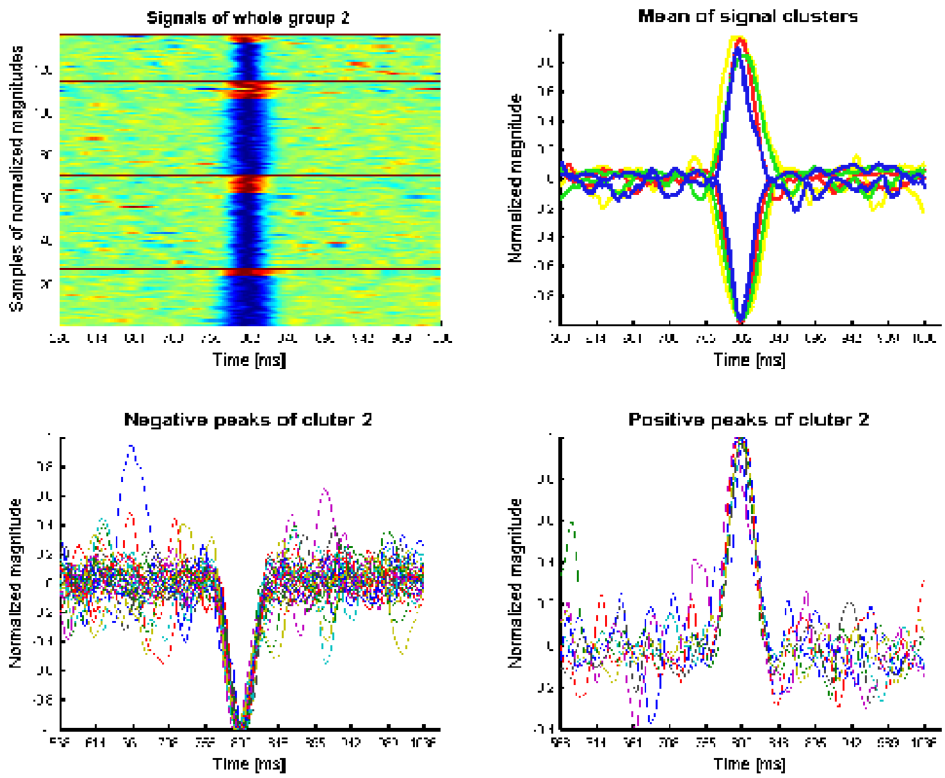

Figure 3.17 Detailed results of one group of real data set. The data of signals, which had peak in second interval are displayed. Signals present in different clusters on the upper left picture are separated by red vertical lines. The upper right picture shows mean values of single clusters. Two lower pictures are showing peakse, separated according to their positivity and negativity.

3.6 Discussion

As we can conclude from previous figures and the program outputs published in appendix section, the algorithm shows good ability in separation and grouping the ERP signals. It proofs, that time frequency analysis can bring interesting results and can be used in the ERP clustering and classification. However, some problems have been found out and it is necessary to note them.

Concerning the artifact rejection phase, the FFT clustering or classification methods have not been found useful in this context. Problematic ERP signal peaks do not have a sinus like shapes and it is very difficult to locate them using FFT. The classification of pre-filtered signal with removed high frequencies demonstrated much better ability in searching artifacts.

The weakest point of the STFT application and frequency features clustering is its sensitivity on the input data. If all the artifacts are not properly removed beforehand, the method shows capability in the separation of artifacts from real signals, but it looses its ability to separate sensitively the peaks, according to their frequencies. The difference is obvious on Figure 3.16 and Figure 3.18, where the algorithm is applied as a whole and as a whole excluding the artifact rejection phase. On contrary, the algorithm shows its interesting features and power in the signal grouping even without the artifact rejection phase. The sensitivity can be improved by eliminative methods, which can be applied on the selected window immediately after it is located. The energy of signal within a window rating can be a good way for possible rejection of unwanted artifacts from the consequent step of clustering.

The possibility of phase centering and phase coherence search is also limited, since some of the ERP signals can consist of more peaks closely following each other, which is impossible to locate exactly by STFT window search method, designed for one peak selection (Figure 3.19). On the other hand, not many different kinds of ERP's are usually present in the dataset and in this case, the method can perform well, since the ERP's can have similar nature and frequency characteristics.

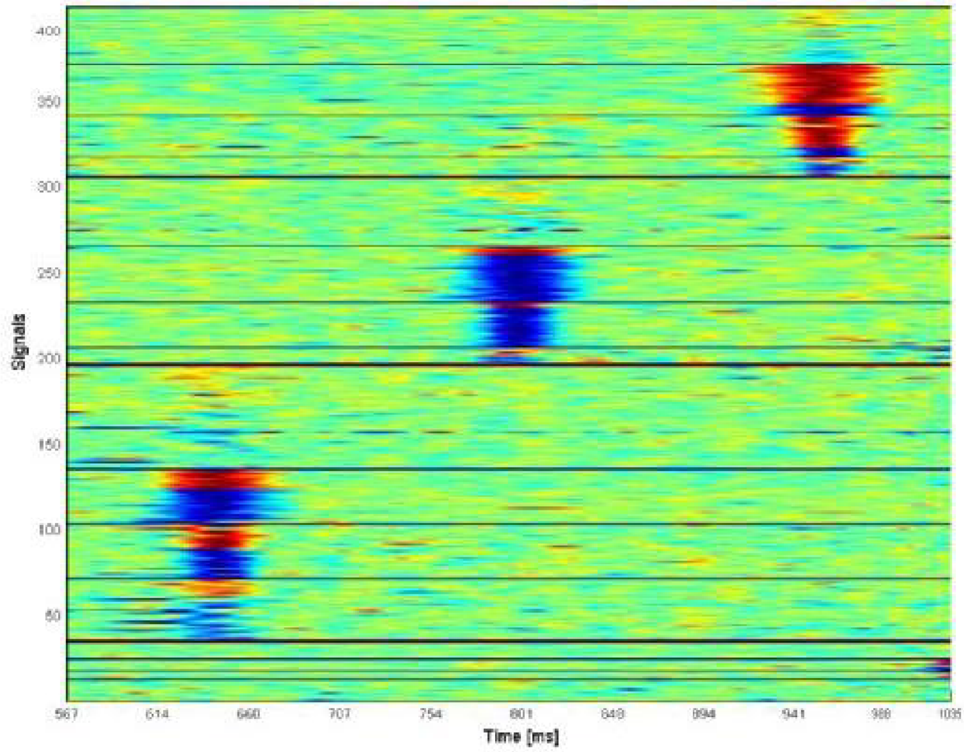

Figure 3.18 Result of the STFT grouping algorithm performed on the data which include all the artifacts. Method was applied on 1/3 of the real dataset. Latency groups are separated by red lines, the frequency feature clusters within the groups are separated by blue lines. Artifacts are present in the same clusters as significant peaks are separated from them.

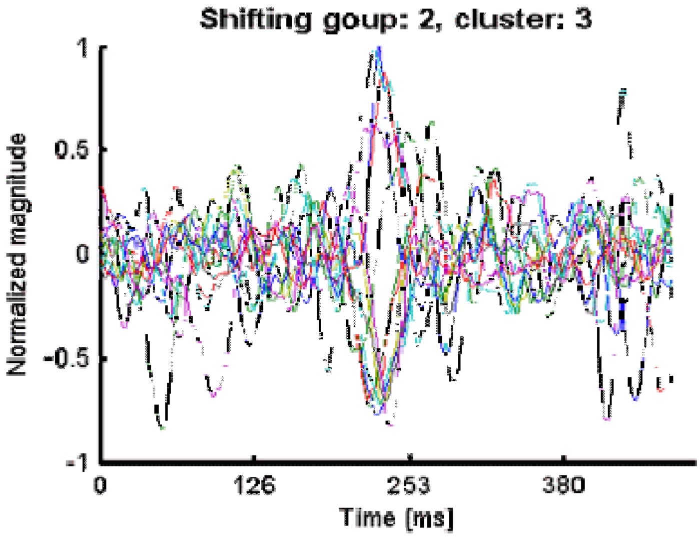

Figure 3.19 Wrongly detected phase coherence of signals within the group.

3.7 Integration to the SLA

The algorithm proposed and presented in this work is a demonstration of the solution using frequency and time frequency approach. It was implemented with the consideration of possible later integration. However, because the final code of SLA was not released in the same time, its final integration with time frequency approach could not be done.

Single sub-steps of both algorithms are corresponding with each other, as it is possible to see on Figure 3.20. Two main layers can be discussed in the context of integration.

The artifact rejection phase is considered as a supplement of original SLA sub-step. It demonstrated high performance and experience had shown that it positively affects the performance of STFT analysis with significant importance.

The original first stage temporal clustering phase can be fully replaced by proposed STFT approach to return similar results or can supplement the current SLA layer. This assumption needs to be supported by the real experiment. Nevertheless, this method offers new features for grouping, which can be found beneficiary, especially during tests on new unknown data. As it was discussed before (section 3.5), the accuracy of main method depends on pre-filtering phase of input data much more than the original one in SLA. This fact with its higher computing performance requirements are the main disadvantages of new approach.

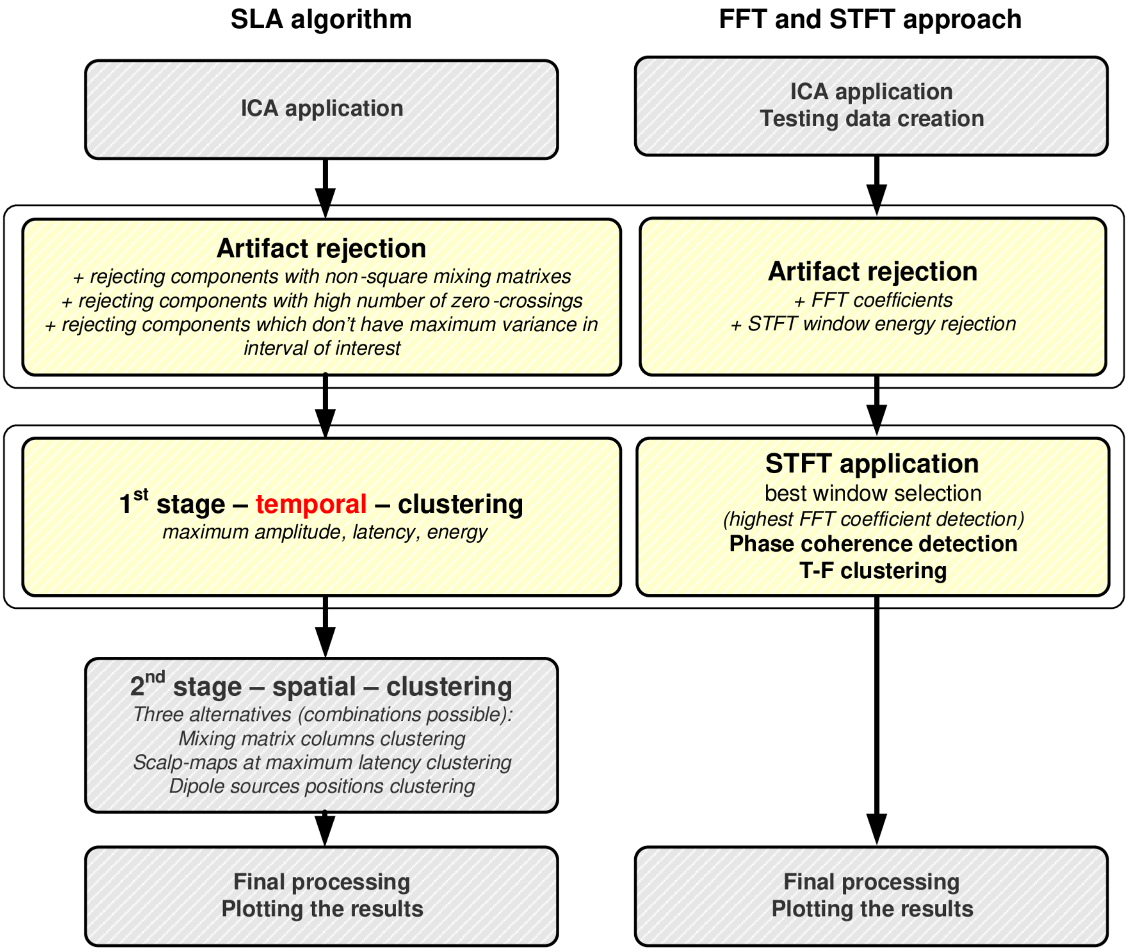

Figure 3.20 Correspondence of original algorithm (SLA) and new approach proposed in this work.

4 Conclusion and future work

The main intent of this work was to investigate new ways of analysis the EEG signals in a context of the ERP classifying algorithm developed at the BSIG labs of TEI. Based on the description and the analysis of the master algorithm and the progress of the solution, three time frequency analysis methods were chosen and their performance was shown and discussed besides the possible way of integration.

Before the actual performing of the methods, a special algorithm for artificial data modification was proposed since the inconvenient frequency properties of the dataset. The original dataset consisted of signals with similar frequency characteristics and was not suitable for time frequency analysis testing. Finally, the method was tested on the real dataset, the positive results were presented and the weaknesses were discussed.

The FFT approach was used to analyze the artifact signals in the EEG data. Although the solution was successful in the context of artificial dataset, the results acquired from real dataset were not satisfactory. Thus, a new approach was introduced. The signal filtered by low-pass FIR filter was analyzed and than classified. The solution has demonstrated performance good enough for the next step: Short-time Fourier Transform analysis. The window with significant ERP like wave is localized by searching the output of STFT analysis. Based on the selected window, the sample is grouped and shifted. According to the master approach results and the requirements, our approach has demonstrated that it can be a good supplement in the master program for ERP grouping. However, the conclusions and results have not been proved by sufficient amount of tests. The most demanding future work can be concluded from this shortage. It is necessary to acquire more data from the partners and other possible institutions to examine and possibly improve the method. Since the ERP signals can significantly vary from subject to subject and can be significantly modified by various disorders, the analysis of these impacts need to be done.

New questions arise from the problem of integration of the current approach to the original SLA solution. The frequency analysis methods can bring a positive contribution to the original concept as it is noted previously, but this assumption needs to be proved by real software integration and analysis of the results. The work especially lacks the discussion of results gained by combination of two stage clustering of the frequency and the spatial features. This could not be done, since the spatial feature clustering stage has not been implemented to the solution. The consultation of clustering of time frequency features in the combination with spatial information of the independent components has not been given due to the lack of time and possibilities of the testing. This fact is probably the weakest point of this work, since the combination of these two approaches can be at current stage unpredictable, which is not desirable.

The other problematic point is the fact, that the approach has not been discussed with neural specialists. Their point of view would surely improve the quality of the project work and show its strengths, weaknesses and other future suggestions. However, this effort is a challenging task and needs a promotion and plenty of partner researchers, which makes it more challenging task for postgraduate thesis work rather than graduate level one.

The methods based on the Fourier transform were chosen as a most common approach for frequency and time frequency analysis. It is a first natural step of investigation in the context of our master algorithm. However, other solutions were not investigated and compared, which could be also desirable. The performance of the proposed solution has not been compared with wavelet transform as well as the other time frequency instruments, such as a complex solution by using filters. One can find many possibilities of how to use the same techniques for mining different features as well as to pick a technique from a wide set of other possible tools for time frequency analysis. This area offers probably the most extensive space of possibilities and suggestions for future work. The concept could be in this way compared in the context of related techniques and the final judgment about the convenience of the method could be than given.

The work on this project made me to investigate basic signal processing approaches and techniques. Moreover, I have found the area of neurology and EEG very attractive and I welcomed it greatly as a topic of my Master thesis project. During the implementation process, I got familiar with MATLAB environment, which I found to be a powerful tool useful and ideal for academic research in the mathematical dimensions.

Finally, I would like to greatly thank to Prof. George Papadourakis and Panagiotis Nikolopoulos at TEI of Crete for their support and generosity, which leaded me in my investigation process and allowed me to acquire a valuable knowledge.

Literature

[1] BIOPATTERN Network of Excellence FP6 programme web portal, http://www.biopattern.org

[2] Brain Special Interest Group of TEI Crete, BIOPATTERN member, web portal, http://biopattern.teiher.gr

[3] David A. Peterson, James N. Knight, Michael J. Kirby, Charles W. Anderson, Michael H. Thaut: *Feature Selection and Blind Source Separation in an EEG-Based Brain-Computer Interface*, EURASIP Journal on Applied Signal Processing 2005:19, pp. 3128–3140, 2005

[4] Julie Onton, Marissa Westerfield, Jeanne Townsend, Scott Makeig: *Imaging Human EEG Dynamics Using Independent Component Analysis*, Neuroscience and Behavioral Reviews 30, pp. 808-822, 2006

[5] A. A. Green, M. Berman, P. Switzer, and M. D. Craig: *A Transformation for Ordering Multispectral Data in Terms of Image Quality with Implications for Noise Removal*, IEEE Trans. Geosci. Remote Sensing, vol. 26, no. 1, pp. 65–74, 1988

[6] A. J. Bell and T. J. Sejnowski: *A Non-linear Information Maximization Algorithm that Performs Blind Separation*, Advances in Neural Information Processing Systems 7, pp. 467–474. The MIT Press, Cambridge, MA, 1995

[7] Jean-Francois Cardoso, Antoine Souloumiac: *Blind Beamforming for Non-gausian Signals*, IEE – Rroceedings – F, vol. 140, no. 6, pp. 362-370, Dec 1993.

[8] A. Hyvärinen and E. Oja: *Independent Component Analysis: Algorithms and Applications*, Neural Networks 13(4-5), pp. 411-430, 2000

[9] Bareš Martin: *Contingent Negative Variation (CNV) – Theoretical Background and Practical Use in Neurosciences and Psychiatry*, Psychiatrie 03/2001, Tigis 2001

[10] Linden David E. J.: *The P300: Where in the brain it is produced and what does it tell us?*, The Neuroscientist, Volume 11, Issue 6, pp. 563-576, 2005

[11] E. Relly: *EEG Recording and Operation of the Apparatus*. In: E. Nidermayer and F. H. Lopes da Silvia (eds). *Electroencephalography: Basic Principles, Clinical Applications and Related Fields*. Baltimore: Williams & Wilkins, 3rd ed., pp. 104-124, 1993

[12] R. Quian Quiroga: *Quantitative Analysis of EEG signals: Time frequency Methods and Chaos theory*, The wavelet digest, Volume 8, Issue 6

[13] H. Berger: *Über das Elektrenkephalogramm des Menchen.* Arch. Psychiat. Nervenkr. 1929;

[14] E. Niedermeyer, *The Normal EEG of the Walking Adult*. In: E. Niedermeyer and F. H. Lopes da Silva (eds). *Electroencephalography: Basic Principles, Clinical Applications and Related Fields*. Baltimore: Williams and Wilkins, 3rd ed., pp: 131-152, 1993

[15] L. Kaufman and P.J. Rousseeuw: *Finding Groups in Data: An Introduction to Cluster Analysis*, Wiley, 1990

[16] R. Quian Quiroga, O. W. Sakowitz, E. Basar, M. Schürmann: *Wavelet Transform in the Analysis of the Frequency Composition of Evoked Potential*, Brain research protocols 8, pp. 16-24, 2001

[17] R. Resnick, D. Halliday: *Physics part I*, John Wiley & Sons, Inc., ISBN 0-471-71717-7

[18] Emmanuel C. Ifeachor, Barrie W. Jervis: *Digital Signal Processing, A Practical Approach, Second Edition*, Prentice Hall 2002, ISBN 0-201-59619-9

[19] Allan V. Oppenheim, Allan S. Willsky, S. Hamid Nawab: *Signals & systems, Second edition*, Prentice Hall 1997, ISBN 0-13-651175-9

[20] Zdeněk Vojtěch: *EEG v Epileptologii dospělých*, Grada 2005, ISBN 80-247-0609-3

[21] German Gomez Herrero, Eero Huupponen, Barrie Jervis: *Blind source separation techniques for processing electroencephalographic recordings* , Tampere University of Technology, BIOPATTERN network of excellence, December 30, 2004

Abbreviations

TEI	Technological and Educational Institute of Crete
BUT	Brno University of Technology
BSIG	Brain Special Interest Group of TEI Crete
EEG	Electroencephalogram
ERP	Event related potential
ERP	Evoke related potential
MEG	Magnetoencephalogram
BSS	Blind source separation
ICA	Independent Component Analysis
SLA	Source localization algorithm (the original algorithm developed by BSIG)
FFT	Fast Fourier Transform
STFT	Short-time Fourier Transform
FIR	Finite impulse response filter

List of Appendices

Appendix 1: Brief description of implementation

The program implementation is based on two toolboxes. The toolboxes are present on the report CD in the directory matlab_toolboxes. The implementation is present in the matlab_implementation folder.

EEGLAB toolbox, version 5.03, is designed for analysis and manipulation of EEG data. It consists also tools for BSS algorithm application and time frequency analysis. The rich user interface provided with the toolbox has not been used during the evolvement.

ICAEP toolbox was provided by the University of Sheffield shortly before finishing this project work. The toolbox is used for ICA application on the raw data. Back projection is performed outside this toolbox.

The program implementation itself is divided in two directory folders named scripts and functions. As the names are expressing, MATLAB scripts, which perform global tasks and integrate smaller pieces of code are placed in the first folder and functions, where the core implementation is done are in the functions folder. In the scripts folder procData01.m and procData02.m are the most important scripts, since there is implemented whole the process, which is described in this work. The first file demonstrates the principles on real data as the second file works on artificial ones. Majority of functions has the prefix, which expresses the nature and way of use of the function (except of projectnew.m, absmax.m and eegplot_jan.m):

- viz - plotting and printing operations
- fio - file output operations
- fft - FFT application and processing
- stft - STFT application and processing
- fir - FIR filter application and classification
- tra – Signal transformation for gaining new samples

Most functions have a commented code with brief explanation of its use in the beginning, which can be easily plotted in the matlab console window by command help.

Additional directories are necessary for manipulation with the data. The inicial-data directory contains raw data in special format as they were received by our partners. The other directories rawData, formattedData and ICAResults are used by ICAEP toolbox for data manipulation. They reflect sub-steps of their algorithm application.

To run the scripts successfully, three configuration files need to be modified: startup.m file in the implementation directory, settingTheInitialParameters.m and stPath.m, both present in the ICAEPtoolbox directory. The directory and toolbox paths need to be set properly in these files.

Appendix 2: Results of real dataset

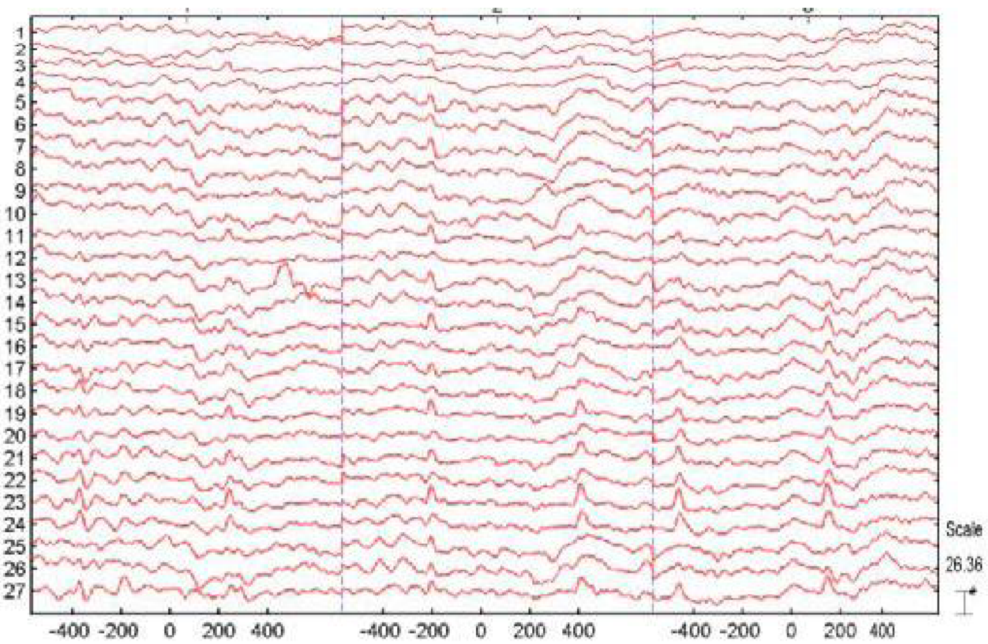

Figure 4.1 Samples of three random epochs of original signals

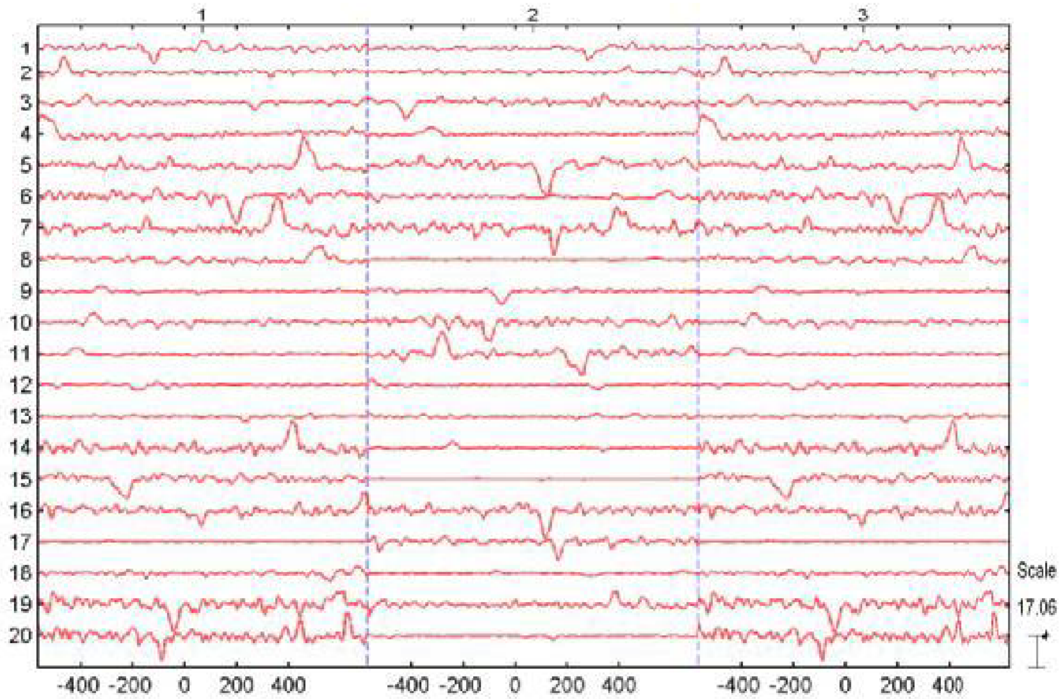

Figure 4.2 Preprocessed ICA components after back projections of three random epochs

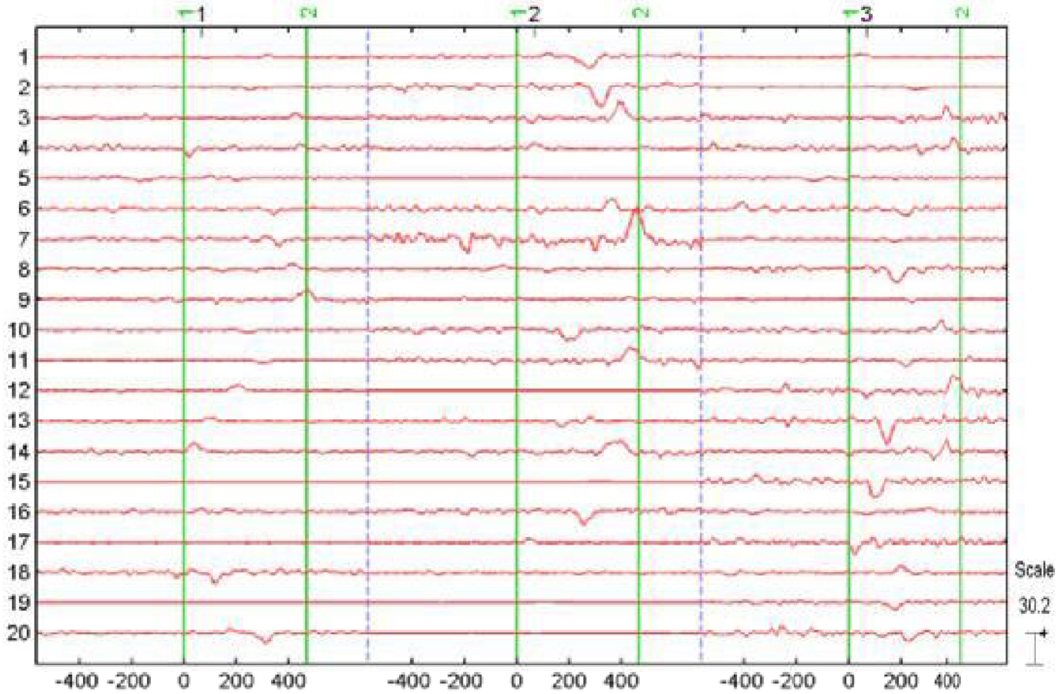

Figure 4.3 Samples that passed through the 1st stage rejection phase (plotted in three columns). Green lines denote interval of interest.

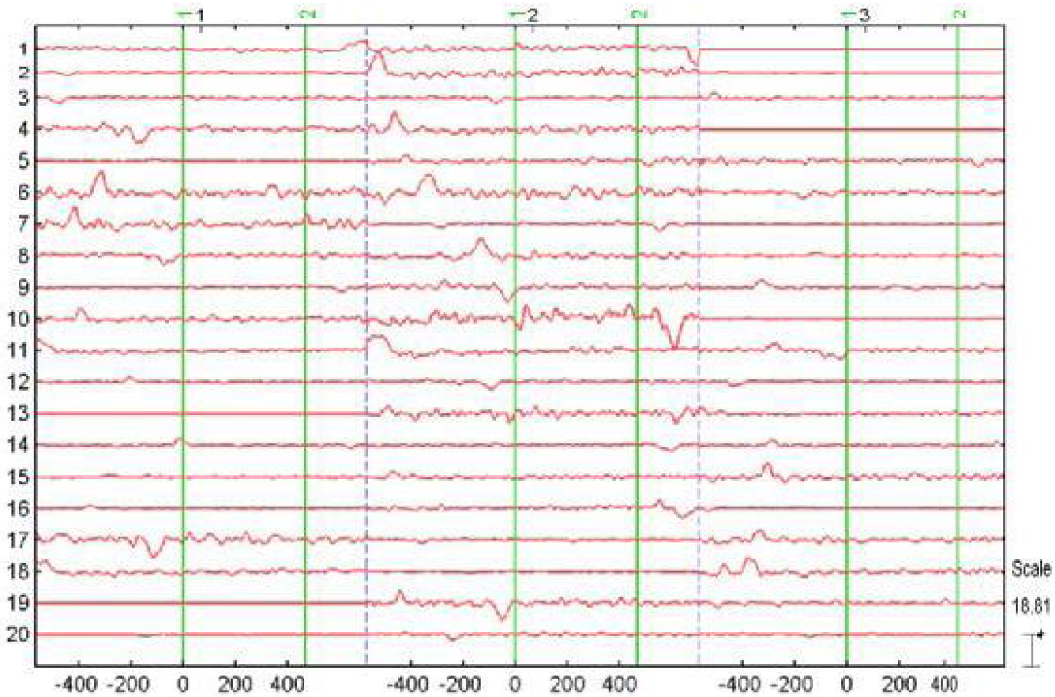

Figure 4.4 Rejected samples of the 1st stage rejection phase (plotted in three columns). Green lines denote interval of interest.

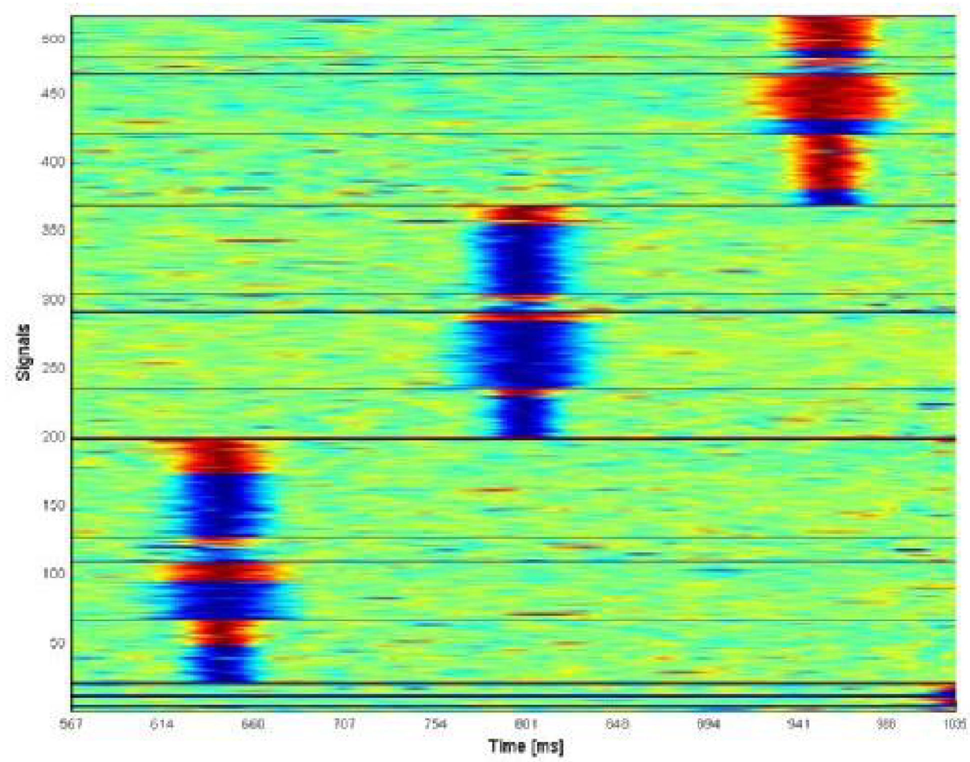

Figure 4.5 Overal grouping result after STFT application. Three significant groups (separated by red horizontal lines) with similar latency consits of four frequency features clusters (separated by blue horizontal lines). The last group in the lowest part of figure containes rejected signals, which peaks exceed the border of interval of interest.

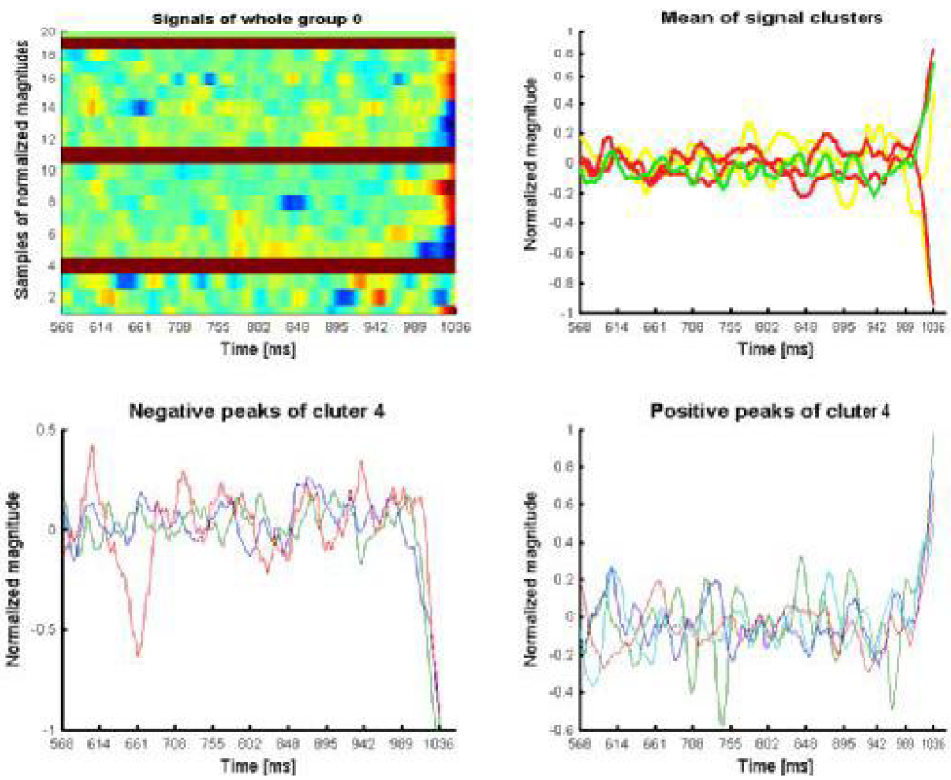

Figure 4.6 Detail of group of rejected signals on the second stage of grouping.

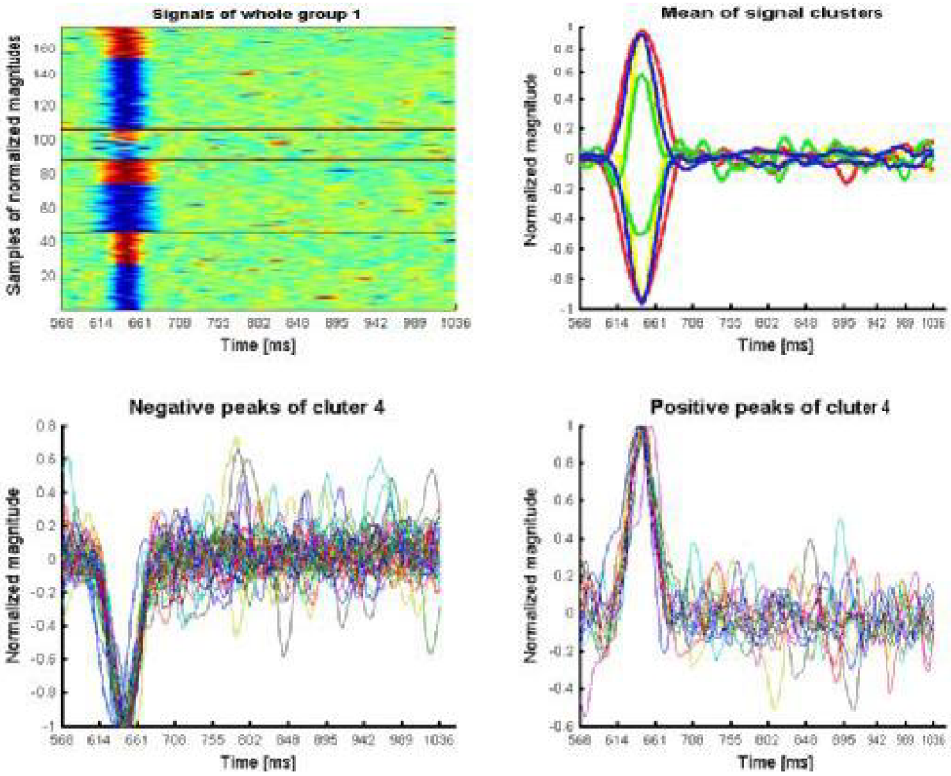

Figure 4.7 Detail of the group of signals with peak around 650ms. Clusters of frequency features are separated by red horizontal lines.

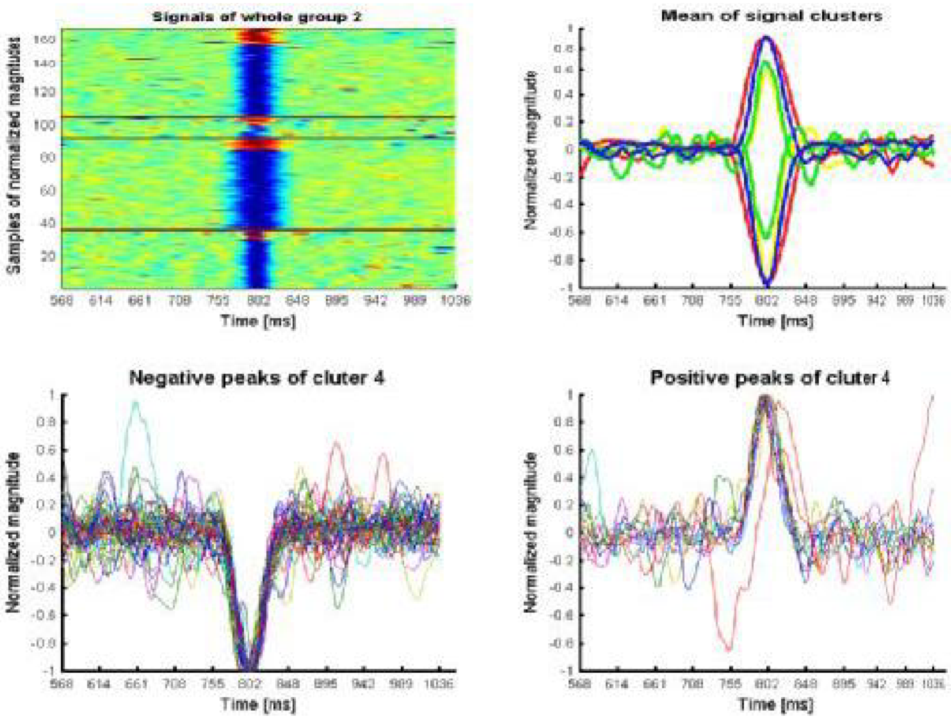

Figure 4.8 Detail of the group of signals with peak around 800ms. Clusters of frequency features are separated by red horizontal lines.

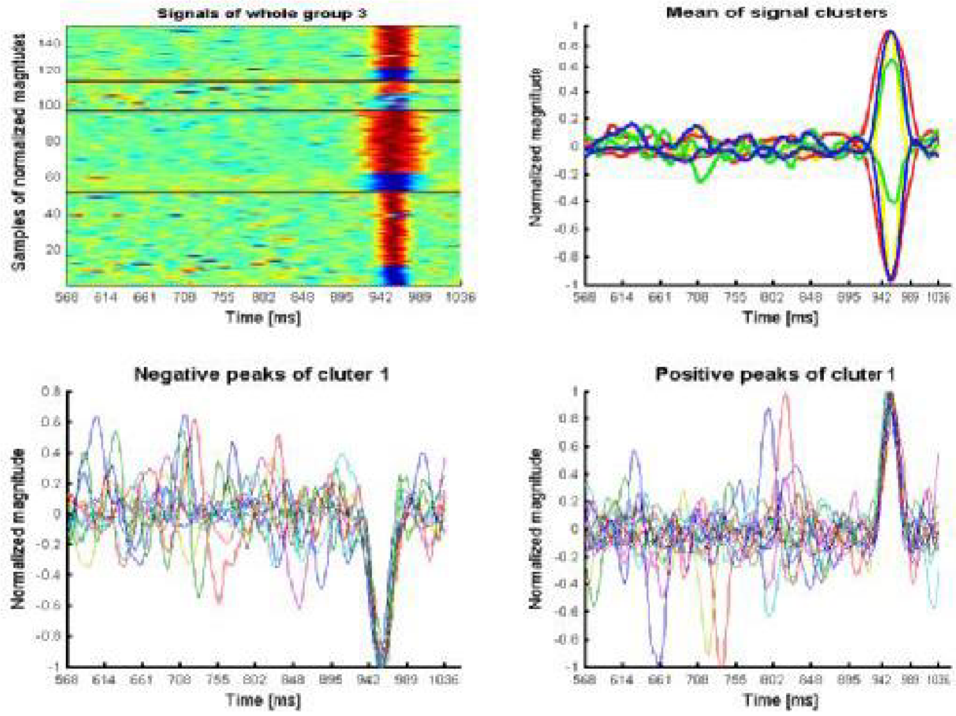

Figure 4.9 Detail of the group of signals with peak around 950ms. Clusters of frequency features are separated by red horizontal lines.

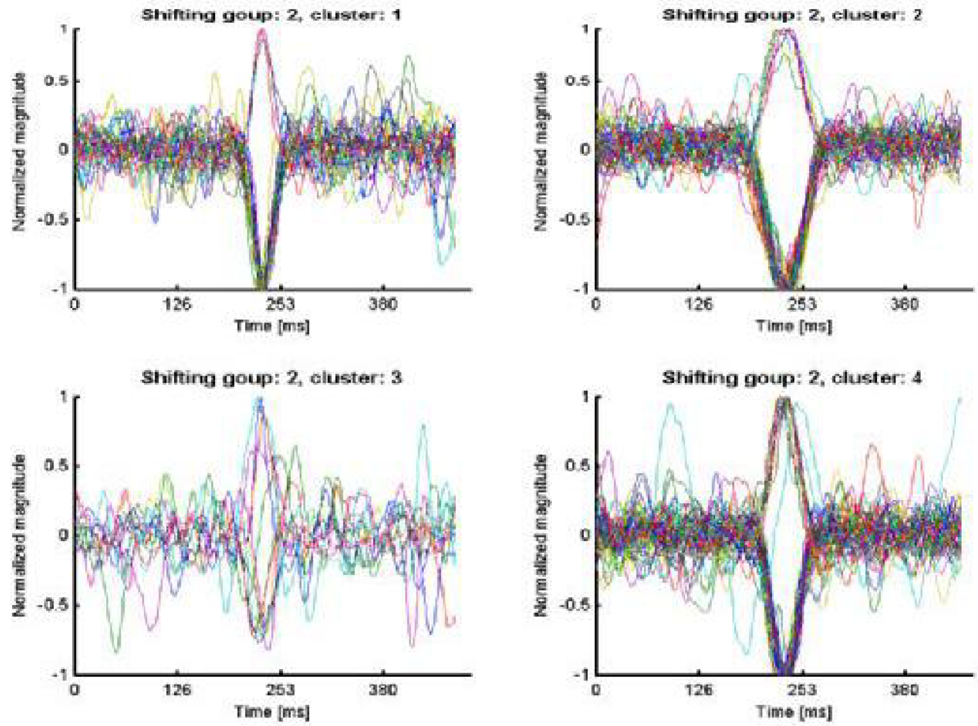

Figure 4.10 All the signals grouped as their peak is in the same latency interval.