

Exploratory data analysis of evoked response single trials based on minimal spanning tree

Nikolaos A. Laskaris, Andreas A. Ioannides*

Laboratory for Human Brain Dynamics, Brain Science Institute, RIKEN, Wako-shi 351-01, Japan

Accepted 15 December 2000

Abstract

Objective: An exploratory data analysis framework, based on minimal spanning tree, is proposed as a means to support the analysis of single trial (ST) electrophysiological signals. The core of this framework is the compact description of the input ST sample in a form of content-dependent ordered lists. Based on the established hierarchies, efficient ways to increase the SNR, extract prototypical responses, visualize possible self-organization trends in the sample and track the course of evoked response along the trial-to-trial dimension, are proposed.

Method: Magnetoencephalographic auditory evoked responses were used for demonstrating and validating the introduced framework.

Results and conclusion: The results demonstrate the benefits, from this intelligent manipulation of STs, in understanding and enhancing the actual evoked signal. Specifically we find support for stimulus-induced phase-resetting hypothesis in the 3–20 Hz band, the existence of trials void of the prototypical evoked response, and an order across the single trial set hinting at an underlying process with long time scale. © 2001 Elsevier Science Ireland Ltd. All rights reserved.

Keywords: Single trial analysis; Cluster analysis; Selective averaging; Visualization; Monitoring; M100

1. Introduction

The study of *evoked potentials/evoked fields* (EPs/EFs) requires separation of the brain's response from the ongoing brain activity. To this end, a great number of *single trials* (STs) are recorded and subsequently averaged to produce a brain's response estimate. The assumption of a stereotypical time-locked signal contaminated with Gaussian zero-mean noise, inherent to this conventional approach, is rarely fulfilled in practice. Many factors, which are difficult to control are unavoidably encountered in the course of an experiment. For example, the overall and moment to moment state of the subject is changing by shifts in attention and fatigue, while habituation and previous history (e.g. contingency of stimulus with the subject's previous history) contribute to a dynamic behavior of the response even during short recording sessions. Moreover, the major interest of brain research currently lies in the analysis of non-stationary brain processes. Cognitive neuroscience introduces experiments involving learning, habituation and memory handling (Liu et al., 1998b; Cansino and Tellez-Alanis, 2000; Rama et al., 2000). For such a research direc-

tion an ensemble-averaging type processing of STs constitutes a 'static' viewpoint which compresses and in some occasions even eliminates dynamic aspects of the evoked response.

The information conveyed by the individual STs, is rarely taken into account. The vast size of ST samples and their poor signal-to-noise ratio (SNR), make the understanding and handling of these samples unattainable. Classification and clustering techniques have been introduced to support their analysis (Gath et al., 1985; Geva and Pratt, 1994; Zouridakis et al., 1997a; Geva, 1998; Lange et al., 2000). The final output is based on patterns representing the individual STs and constitutes a categorization of these patterns into homogeneous classes. Each one (or even more) of these classes may reflect a brain behavior (e.g. spontaneous reaction, anticipation) stereotypical for a specific type of stimulus.

While *parametric* unsupervised/supervised classification schemes could guarantee optimality, at least in a statistical sense, the statistical assumptions about the number of classes, a priori, probabilities of the classes and the probability density distributions within the classes which must be satisfied make these schemes inapplicable in practice. *Nonparametric* schemes do not imply models for the underlying distributions and therefore could deal with the statis-

* Corresponding author. Tel.: +81-48-467-9730; fax: +81-48-467-9731.
E-mail address: ioannides@postman.riken.go.jp (A.A. Ioannides).

tically unpredictable nature of STs. However, they have an error-prone automatic character. Basic parameters, like the number of expected classes, are either required to be given before execution and used without being validated or estimated during execution based on heuristics not fully compatible with the analysis of STs. To resolve this situation, *exploratory data analysis* (EDA) can be employed. Under the general term EDA a broad collection of mainly non-parametric techniques is included, ranging from *projection pursuit algorithms* to *directed searches* and clustering algorithms and sharing as common objective the understanding of the input sample.

Here, an EDA framework is suggested as a means to gain informative insights to the ST sample. In this framework a p -dimensional ‘*features-vector*’ is extracted from the *pattern vector* (i.e. the signal values in the time domain) of each ST. In this way each ST is considered as a point X_i in a multidimensional space \mathbf{R}^p and consequently the set of STs can be treated as a collection of points in \mathbf{R}^p . Geometric ideas can then be used leading to an effective description of adjacency relationships among individual STs. The intuitive idea behind this consideration is that attractive (repulsive) forces are mutually applied between points whenever the associated STs contain the same (different) EP/EF component. Hence, densely populated neighborhoods in \mathbf{R}^p correspond to subset of STs containing the same well defined EP/EF component. By the same token, the ongoing activity is responsible for the attenuation of these forces, smearing the borders between neighborhoods of different EP/EF component.

The backbone of the proposed point set description is the *minimal spanning tree* (MST), used by many graph-theoretic clustering approaches (Jain and Dubes, 1988; Chowdhury and Murthy, 1997). Here, MST captures the cohesion between STs and enables description of the input sample in the form of *ordered lists*. The proposed ordering scheme, named hereafter *MST-ordering*, facilitates intelligent queries in the ST sample and can act as *dominant-mode seeking clustering* algorithm. In this way subsets of STs with (same) high quality EP/EF content can be delineated.

Emphasis was put into the visualization of the MST-ordering procedure. An efficient technique known as *MST-planing* is revisited (Lee et al., 1977). It draws the MST on a plane inducing minimal deformation. This effective shrinkage of the original p -dimensional space to a two-dimensional one enables visual exploration. The great interest in such a technique comes from the emerging technologies, which facilitate the interaction of man and machine. With the user’s interaction, supported by -nowadays- trivial *Graphical user interface* (GUI) tools, many complicated or laborious tasks of ST analysis, like artifact detection, can be simplified.

The *multivariate Wald-Wolfowitz test* for intersample comparisons accompanies the methodology, enabling comparisons between different recording conditions (e.g. spontaneous vs. evoked activity, start vs. end of the record-

ing session) and serving as a guideline for selecting the EP/EF signal representation.

Magnetoencephalographic data from simple auditory experiments were used to demonstrate/validate the proposed EDA framework. The 151 whole cortex MEG-signal offered a unique environment for validating the core of the proposed framework, the MST-ordering. The *virtual signal* (VS) transform was applied in order to amplify the contribution from the complex of generators at or near the auditory cortex generating the dominant M100 response. The VS output is practically free of disturbances like oculographic or MCG artifact (Liu and Ioannides, 1996; Liu et al., 1998a). Based on this enhanced signal a chain of consecutive values of magnetic field strength was used as the pattern representing the M100 component in each ST. Using this simple representation it is shown that a hierarchy of nested subsets can always be produced, in a way that the SNR of the STs progressively deteriorates. Different subsets of STs can be delineated: subsets with strong well time-locked response and subsets with no response time locked to the stimulus onset or with response buried in the ongoing activity. The existence of such subsets appears to be a common feature of real electrophysiological data. The ability of MST-based procedures to reliably reveal the structure of ST samples is further demonstrated via a simple classification task and via a technique to monitor the SNR. The extension of the proposed methodology, so as to work on truly multichannel data and possible applications (e.g. *intrahemispheric binding*), although straightforward, is treated only as part of the discussion.

The paper has been divided into two parts. Section 2 serves as an introduction to the proposed EDA framework. Section 3 is devoted to the experimental results. However, for the clarity of presentation, a few instructive results have been included also in Section 2. Technical details which are necessary for understanding the details of the proposed algorithms have been given in the appendices.

2. Theoretical framework

2.1. MST

Graph theory sketches the MST structure with the following definitions. A *graph* is a set of nodes and a set of node pairs called edges. An *edge-weighted* graph is a graph with a real number, called weight, assigned to each edge. A *connected graph* has a path between any two distinct nodes. A *spanning tree* is a connected graph that includes all the nodes without loops. The MST is the spanning tree of minimum total weight.

When the previous concepts are applied to a set of N points, a node is dedicated to each point and the corresponding pairwise distances (or generalized dissimilarities) are assigned as weights to the formed edges. The MST is the

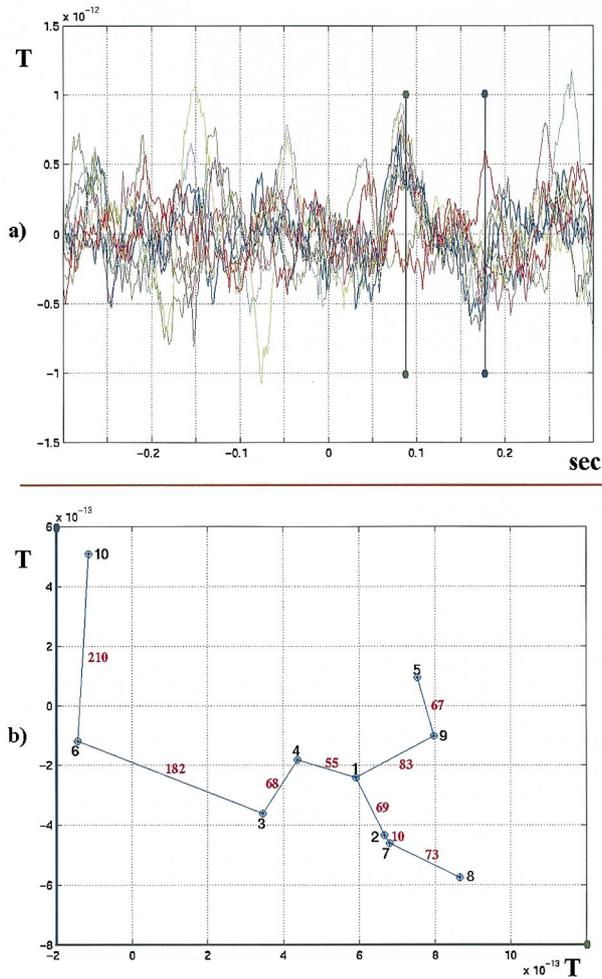


Fig. 1. (a) A sample of 10 STs, with the two vertical lines depicting the axes used to represent the STs as points in \mathbf{R}^2 . (b) the MST of the corresponding point-sample.

connected graph, emerged from the collection of exactly $(N - 1)$ edges, having minimum total length.

In order to demonstrate how the previous abstract graph-theoretic concepts are used to handle STs, a simplified two-dimensional example, from real data, is given in Fig. 1. With each one of the 10 STs shown in Fig. 1a, a point in \mathbf{R}^2 is associated. The two-dimensional configuration of this point sample is given in Fig. 1b. Each of the two axes spanning this reduced space expresses the strength of the magnetic field at a time instant in the poststimulus range. The first (horizontal) axis was selected, by visual inspection, so as to correspond approximately to a time instant where the majority of STs presents a positive deflection. The second (vertical) axis corresponds to a time instant chosen at random. In this graph the 10 points appear as nodes indexed from 1 to 10. These indices reflect the physical time order of the corresponding STs. The MST appears as a collection of 9 line segments, the *edges*, with sample points as endpoints. The *weight* associated with each edge is also indicated. It is the pairwise Euclidean distance between the corresponding

points. A scaling has been applied on these distances such that the smallest of them (2.991×10^{-14} T) appears on the graph as 10. With such a graph, it is easy to conceptualize the notion of *central/prototypical* points and *outliers*. The term *degree* of a node is used to denote the number of edges incident on it. *Central* points (e.g. 1) differ from outliers (e.g. 10) in terms of degree and weights of the associated edges.

The most straightforward way actually to apply this graph theoretic approach, is the generalization of the above mentioned axis selection, so as to include \mathbf{p} time instants around the approximate latency where the EP/EF component peaks. An estimate of this latency could be provided through the peak latency in the waveform of the averaged STs. Hereafter this simple approach is adopted, so as to configure the STs in \mathbf{R}^p . However, any other technique for the extraction of \mathbf{p} features from the ST time waveforms could be applied. The method presented in Section 2.5 can be employed to select among possible alternative ST representations.

The MST provides a compact description of the point set (Hartigan, 1975). It contains the ‘nearest neighbor’ information about each point and the ‘shortest linkage’ information about subsets of points. There are many heuristic strategies, exploiting these properties, that could be used for identifying clusters in the ST sample (Zahn, 1977). The use of these strategies is suggested only after the user has gained some experience about what information is encoded in this graph structure and how this information is distributed in the different tree branches. The directed graph search and the MST-planing, presented in Sections 2.2 and 2.4, respectively, both contribute to this direction.

2.2. MST-Ordering

Ordering in \mathbf{R}^p has been introduced as a method to study the distribution of multidimensional point sets (Barnett, 1976). It was sought as a generalization of the ordering procedure for real numbers, i.e. points in \mathbf{R}^1 . A complete generalization is not feasible. However, under certain circumstances, some ordering schemes can achieve the basic concept that adjacent points should have similar ranks (Astola et al., 1990; Hardie and Arce, 1991).

The observation that the MST of points in \mathbf{R}^1 connects the points in sorted order, led to the consideration that the MST provides a unique vehicle for the credible ordering of points in \mathbf{R}^p (Friedman and Rafsky, 1979). The traversal of the MST is used to assign ranks to the points. To denote the rank of each point, the following convention is adopted: while ‘ \mathbf{j} ’ ($j = 1, 2, \dots, N$) is the original index of a point (i.e. the time identity of the corresponding ST), its rank indicator is ‘ \mathbf{i} ’ ($i = 1, 2, \dots, N$). After *rooting* the MST arbitrarily, the rank of each point is the order in which it is visited in a *breadth first search* of the tree. This traversal procedure defines a *parent-daughter* relationship, where the parent of $X_{[i]}$ denoted $P(X_{[i]})$ is the immediate *predecessor* of $X_{[i]}$ along the path to the root. The corresponding edge weight, denoted as $e(X_{[i]}) = (\|P(X_{[i]}) - X_{[i]}\|)$, quantifies

(inversely) this relationship. These concepts are demonstrated for the case of the two-dimensional example, shown in Fig. 1. The ordered points $[X_{[1]}, X_{[2]}, \dots, X_{[10]}]$ will be [1,4,2,9,3,7,5,6,8,10] and [8,7,2,1,4,9,3,5,6,10]. If correspondingly X_1 & X_8 is used as the root of the tree. Likewise, the corresponding $P(X_{[i]})$ & $e(X_{[i]})$ will be [1,1,1,1,4,2,9,3,7,6] & [0, 55, 69, 83, 68, 10, 67, 182, 73, 210] and [1,8,7,2,1,1,4,9,3,6] & [0, 73, 10, 69, 55, 83, 68, 67, 182, 210]

Given a sample of STs, the MST-ordering does not only provide a means to navigate – in a systematic manner – into the sample, but also gives rise to signal processing strategies for the recognition of modes in their distribution and the detection of distinct STs. In the simplest case that the corresponding points appear to follow a *unimodal* distribution, rooting at a central point can be used so as outliers to be *unmasked* via the ordering procedure, i.e. to be isolated at the higher ranks; a final test for the rejection of a possible outlier, can be based upon the distance to its parent (e.g. points 6,10 in Fig. 1). To cope with the general case of points following a *multimodal* distribution, an iterative scheme can be adopted. Rooting at a point prototypical for the most significant mode, the ordering enables the delineation of the subset of points belonging to this mode, by locating them in the lower part of the ranked list. The detected subset of points is removed and the procedure is repeated for the rest of the points.

A critical issue in the MST-ordering is the selection of root points. An automated procedure for this selection, based on the *local point density*, is presented in the next section. As a second alternative the user's assessment, facilitated via the dimensionality reduction technique presented in Section 2.4, is also proposed.

It is easily deduced that the MST-ordering is a context based search which can be incorporated into a *subtractive* type clustering scheme (Yager and Filev, 1994; Sindoukas et al., 1997), that can function in either an automatic or semiautomatic mode. In the former mode the root selection and the *cluster validation* are internally controlled by heuristics. In the latter mode user's supervision is engaged.

2.3. MST and point density estimators

The concept of *local point density* (**PD**) has been extensively studied in pattern recognition literature, giving rise to *mode-seeking clustering* algorithms, nonparametric probability density function estimators and *prototypical point detection* strategies (Jain and Dubes, 1988; Chaudhury et al., 1996). The search for global maximum in **PD** has been formulated as an optimization problem for recovering the EP/EF signal from the ongoing activity (Laskaris et al., 1996b). Since points are gathered together according to their EP/EF content, the **PD** measure can act as a nonparametric SNR estimator. After recognizing a critical level \mathbf{PD}_0 corresponding to a random point configuration, SNR increases monotonically with the **PD** measure.

An efficient **PD** estimator, strongly related to the MST, is presented in Appendix A. In the proposed framework, this estimator is utilized in two ways. First, to select the root point as the point of highest **PD** and (optionally) to exclude 'extreme' outliers from further analysis as points of zeroed **PD**. Second, to monitor the SNR during the recording session and, by compensating for changes due to background activity, to track changes in the actual EP/EF signal.

2.4. MST-planing

MST-planing is an efficient technique for mapping a point set $\{X_i\}_{i=1:N}$ on a plane. The basic idea is that with the *sequential mapping* of points originally given in \mathbf{R}^p , some of the interpoint distances can be exactly conserved in the new space (i.e. the plane). Since every triplet of points $\{X_a, X_b, X_c\}$ forms a triangle in the original space, it can always be placed on a plane so that the sides of this triangle are preserved. Having already mapped X_a & X_b onto a plane as point χ_a & χ_b , the image of X_c is defined by the intersection of two circles centered at χ_a & χ_b and having radii equal to the distances of X_c to X_a & X_b respectively. Similarly, the image χ_d of a fourth point X_d can be found by preserving its distance to two of the points in the triplet. Applying this strategy sequentially a map, where $2N - 3$ (out of the $N(N - 1)/2$ in total) distances are unaltered, is obtained. A proper selection of the mapping sequence, can result in a map reliably reflecting the structure of points in the original space. It was suggested (Lee et al., 1977), that this sequence could be provided via the *MST-ordering*.

The algorithmic steps of the above planing procedure are included in Appendix B, where it can easily be seen that this procedure preserves the edges of the MST, hence the term *MST-planing*, offering a credible two-dimensional representation: the original MST is deformed only in terms of multi-dimensional angles between the edges. *MST-planing* preserves also the distance of each point to the selected root. By changing the root of MST, different insights to the original data structure can be easily obtained. Trivial GUI tools can enable this selection via interaction with the current map, and the user can finally select the most 'interesting' one.

The computational simplicity of the *MST-planing* has been underlined by Friedman & Rafsky (Friedman and Rafsky, 1981), who comment that whenever the MST is rooted at a point close to the geometric center of the sample, the resulting map is very similar to the one produced by the application of *multidimensional scaling algorithms* (Kruskal, 1964; Wackermann and Matousek, 1998). Taking this into account, they suggested the normalized total discrepancy as a measure of map credibility, regarding the overall sample structure:

$$E = \frac{\sum_{i < j} |\sqrt{D_{ij}} - \sqrt{d_{ij}}|}{\sum_{i < j} \sqrt{D_{ij}}} \quad \text{where } D_{ij} = \|X_i - X_j\|^2 \text{ and } d_{ij} = \|\chi_i - \chi_j\|^2 \quad (1)$$

Throughout this paper ‘global’ maps are shown, where the point closest to the overall average is used as root, and the E gauge is depicted at top. Such a ‘global’ map, with the MST drawn on it, can be very informative by suggesting paths in the ST sample continuum, along which the signal changes progressively. Abrupt changes may indicate change in the signal content or drastic content deterioration. This is demonstrated via Fig. 2, where after configuring 50 STs in \mathbf{R}^P , the original MST has been planed as given in the left top corner. The time-waveforms of the 50 STs have been also placed on a plane, following the MST-planned configuration of their representative points. In the right top corner the ensemble average of the STs is given, providing the common time/amplitude scale. A *lack of homogeneity in the sample distribution is apparent*. STs of high quality EF content – as this is reflected by well defined peaks of right timing – are gathered at the left. The rest of the STs appear randomly distributed in the opposite side of the plane. A selection of the high SNR subset can even be performed manually, by drawing a line on this plane. Alternatively, the MST can be rooted at a point from the densely populated part of the plane and the MST-ordering will

gather them at the lower part of the list. In a more flexible interactive environment, the signal characteristics (e.g. pre-stimulus activity, ratio of energy in different bands) can be systematically studied by comparing the STs from basins of signal attraction with STs from sparsely populated neighborhoods.

The above cartography of STs can resolve many algorithmic issues in MST-ordering based signal processing (where critical parameters have to be set heuristically), like the number of epochs which should be eliminated because of artifacts and the number of epochs that should be included in a certain cluster. This property arises from the close relation of the MST-ordering and the associated map, which acts in a zooming mode around the selected root.

2.5. The Wald-Wolfowitz test for intersample comparisons

An important branch of ST analysis involves comparisons between different samples of STs, e.g. samples of responses to the same stimulus being applied during different stages of arousal. Moreover, such comparisons can be employed in the designing/training session of signal processing algo-

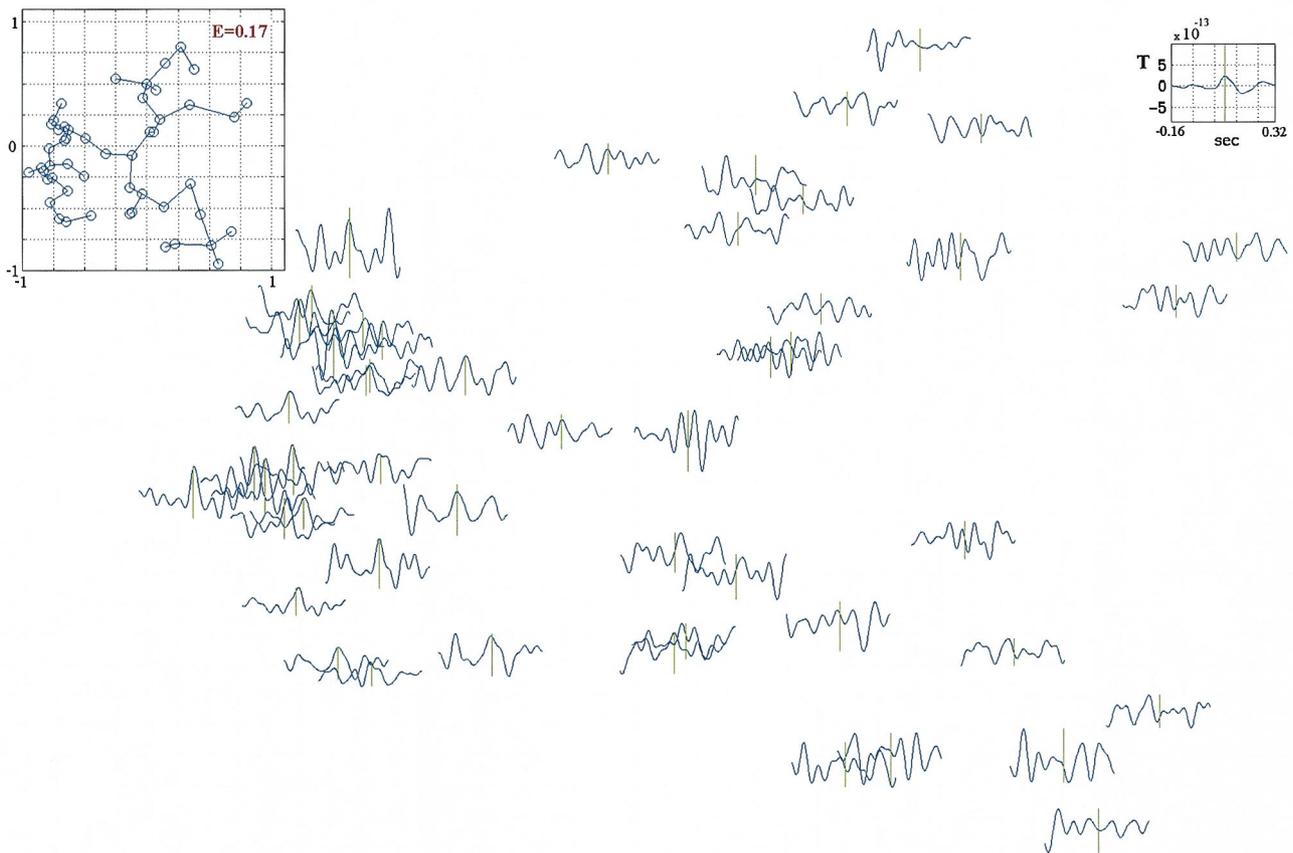


Fig. 2. A sample of 50 STs from binaural stimulation has been configured in \mathbf{R}^P (as is described in Sections 3.2 and 3.3) and the MST has been constructed. In the left top corner the planed MST is given. In the center, the corresponding 50 time-waveforms have been planed accordingly. In the right top corner, the averaged waveform of the ST sample is plotted on the same scale with the individual ST waveforms. Notice the difference in amplitude between the ST waveforms and the ensemble average waveform. The 0 in the time axis denotes the stimulus onset time and the vertical green line indicates (in all waveforms) the latency of the peak in the ensemble average.

rhythms, in order to select/optimize feature extraction procedures. The quantitative expression of intersample ST comparisons must cope with the dynamic nature of STs which incorporates temporal ST contiguity; this is still an open, important but unresolved issue.

As an important component of the introduced framework, a nonparametric test dealing with the ‘*multivariate two-sample problem*’ is suggested. It is a multivariate extension of the *Wald-Wolfowitz test*, comparing two different sets of points in \mathbf{R}^P by checking whether they form different branches in the overall MST (Friedman and Rafsky, 1979). The output of this test, which makes use of the indices \mathbf{R} and \mathbf{W} as described in Appendix C, can be expressed as the probability that the two point-samples are coming from the same distribution. The great advantage of

this test is that no a-priori assumption, about the distribution of the two samples, is needed.

The *Wald-Wolfowitz test*, combined with the MST-planing procedure, offers a unique environment for contrasting different recording conditions or signal representations and may significantly contribute to the understanding of EP/EF signal. An example is presented in Fig. 3 where two ST samples, from different recording conditions, are compared. The first sample, A, consists of 50 STs containing the evoked responses to binaural auditory stimulus and having been collected with certain recording parameters (sampling rate, frequency range, interstimulus interval, etc). The second sample, B, consists of 50 STs having been collected (within the same recording session) with exactly the same parameters, but with the stimulus being disabled. Both

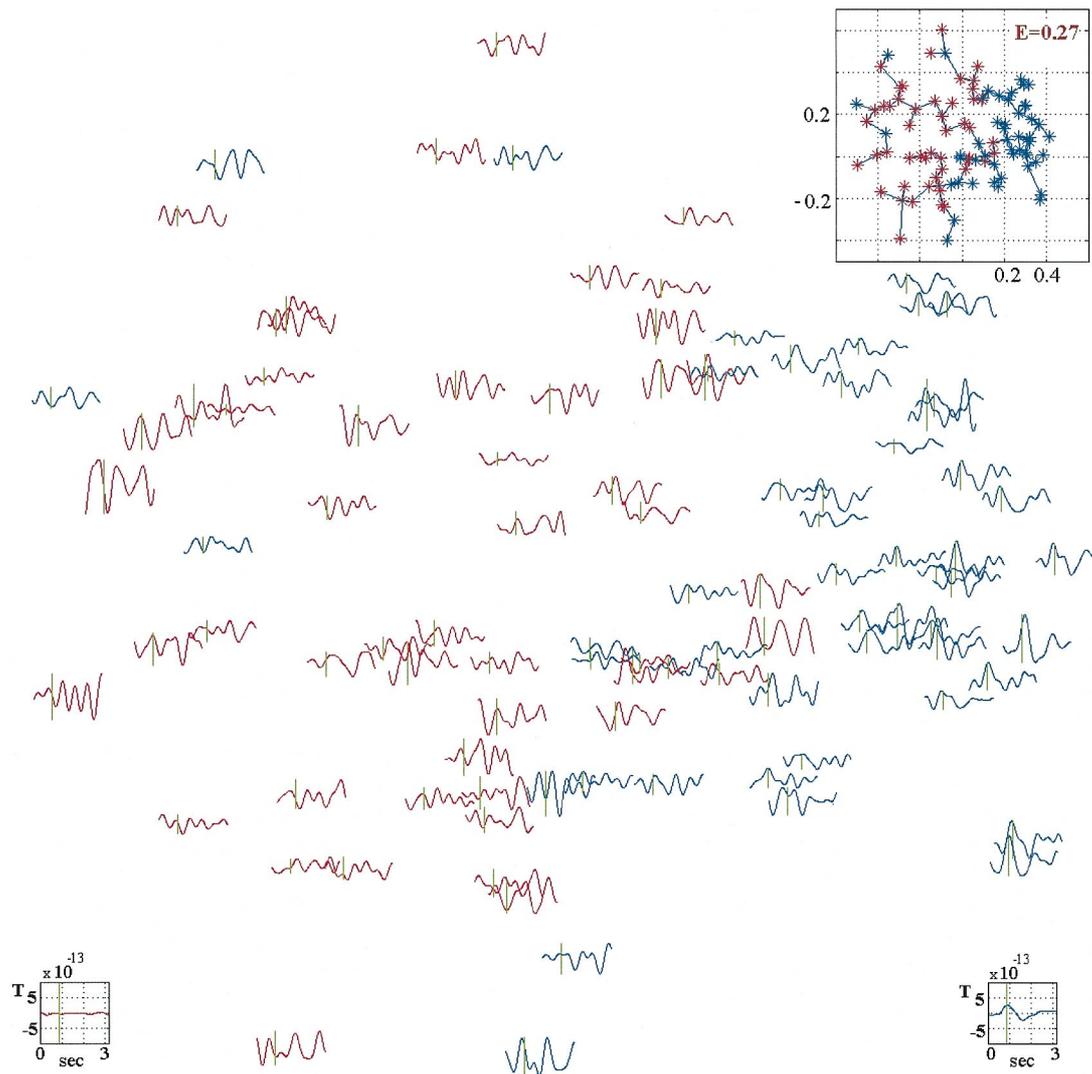


Fig. 3. Intersample comparison of STs: the first sample (denoted by blue color) contains evoked responses to binaural stimulation, while the second (denoted by red color) contains epochs of spontaneous brain activity. Each ST from both samples has been configured in \mathbf{R}^P based on a pattern of $p = 101$ samples extracted around the latency at which the averaged response is maximized (see Sections 3.2 and 3.3). The aggregate MST has been constructed and planed and is shown in the right top corner. The ST time-waveforms have been planed accordingly, with the vertical green line indicating the same time instant. The averaged evoked response (spontaneous activity) is given in right (left) bottom corner.

samples have been configured as points $\{X_i, Y_i\}_{i=1:50}$ in \mathbf{R}^p , based on a representation suggested by the average of A-sample STs which is shown in the right bottom corner. The aggregate MST has been constructed and its planed image is depicted in the right top panel, with the nodes corresponding to the image of the first sample drawn in blue. It can easily be seen that the points corresponding to A and B samples constitute, roughly, different subtrees of the overall MST. The measured \mathbf{R} is 19, resulting in a very small value $\mathbf{W} = -6.429$ (P value = 6×10^{-11}), which clearly indicates the difference in the distribution governing the two samples. The waveforms of the corresponding STs have also been placed on a plane, following the two-dimensional configuration of their representative points.

This comparative study of evoked vs. spontaneous activity, or prestimulus vs. poststimulus activity, can be utilized so as to optimize signal processing parameters (e.g. to define the latency and range p for constructing the pattern of each ST), or to select among different possible representations (e.g. among bands from multirate filtering schemes). The lower the \mathbf{W} , the greater the enhancement in SNR through a certain representation. The interpretation of the corresponding composite map is direct. Additionally, this map suggests a navigation in the overall sample, along directions of steepest change from ‘signal’ region to ‘noise’ region, that can be very instructive regarding the interpretation of ‘evoked response’ and ‘ongoing brain activity’, in each experimental setting.

3. Applications to auditory M100 evoked responses

3.1. Experimental set-up

The CTF whole head system (151 channels) was used to record MEG signals from 3 healthy right-handed volunteers (A.F., V.P., A.O.; two males, one female; age range 25–29 years, mean 27.5). Measurements were carried out inside a magnetically shielded room with the subject sitting in a comfortable position and keeping his eyes open during the recordings. The auditory stimuli were tones (1 kHz, 200 ms duration, 10 ms rise/decay time) presented in all experiments binaurally at constant interstimulus intervals (ISI: 3 s), apart from one additional case (applied to V.P.), where they were presented, in a randomized order, to the left or right ear at random ISI (3 ± 0.2 s). The loudness had been adjusted to a comfortable level for each subject, prior to the recording session (mean level: 45 dB). In each session 120 or 200 stimuli were applied. In all cases, the subject had been instructed to listen passively to the tones.

The MEG signal was recorded in continuous mode, after low-pass filtering at 208 Hz and digitization at 625 Hz. The data were further digitally band-pass filtered in the 3–20 Hz range, using 4th-order zero-phase IIR Butterworth filter, since this range has been recognized as the most important for the quantification of variability in the M100 response

(Liu et al., 1998a). The continuous multichannel signal was segmented into trials lasting from -200 ms to 500 ms, in respect to the stimulus onset.

3.2. The virtual signal transform

The VS transform is a dimensionality reduction technique which computes a linear combination of signals from neighboring channels. It usually involves the difference of two weighted sums of signals, each from one patch of sensors. A data driven method is employed for estimating the weights involved in this spatial filtering action. The output of the VS transform is a signal reflecting mostly the activation of focal generators located below the area between the two patches of channels. The VS transform is an approximation of the smoothed version of an operator involving spatial first order partial derivatives of the signal which produces a quasi-image of the current density below the sensor array (Ioannides et al., 1990). A detailed description of the VS construction for the study of M100 response is reported in (Liu et al., 1998a). They showed that for auditory responses the virtual signal is a good approximation of the activation curve obtained using *Magnetic field Tomography (MFT)* (Ioannides, 1990; Ioannides et al., 1990) to estimate the (non-silent) current density vector from a *region of interest (ROI)* containing the auditory cortex.

In this work, two virtual signals have been constructed, each one dedicated to one hemisphere and named, hereafter, as *left VS* (VS_l) and *right VS* (VS_r) correspondingly. Since we focus on the M100 response, which is an easily recognized early peak dominating the averaged auditory evoked magnetic field, the virtual signals are derived with the following equation:

$$VS_k = \frac{1}{5} \sum_{i_k=1}^5 S_{i_k}^+(t) - \frac{1}{5} \sum_{i_k=1}^5 S_{i_k}^-(t), \quad k = l, r \quad (2)$$

where $S_{i_k}^+$ ($S_{i_k}^-$) denotes the signals from channels which are close to the corresponding auditory cortex and show the greater positive (negative) deflection in the average M100 response.

In the sequel, these two virtual signals are treated independently.

3.3. Signal representation

A data-driven strategy is adopted for the extraction of a pattern conveying the signal content of each trial. First the latency of maximum positive deflection in the ensemble average is identified and the sequence of $p = 101$ timepoints around this latency is selected as the set of latencies of interest. The signal values at the corresponding timepoints of the i th ST are defining the pattern X_i to be used for this trial to be embedded in a p -dimensional space.

This representation endows the subsequent computation of *Euclidean interpoint distances* with total dependence upon the amplitude differences, in the selected latency

range, between the individual STs. Since it has been recognized earlier (Liu et al., 1998a), that the signal amplitude distribution does not show any significant alteration associated with the arrival of the stimulus, the previous representation is not expected to enhance the EP/EF content of each trial. As a potential alternative, the *normalization* of the above pattern is suggested. In this case, the features-vector is denoted as $X_i^* = X_i / \|X_i\|$. The subsequent computation of Euclidean distances measures dissimilarities in shape between the STs. The concept of shape similarity is known to express synchronization and therefore aligns with the admission that the brain's response is encoded in the phase of the individual STs.

Using these two signal representations, the introduced EDA framework enabled the systematic comparison between the two alternative hypotheses: 'amplitude' vs. 'phase' encoding of the brain's responses. STs from the binaural stimulation paradigm were used. Apart from the pattern X_i , a pattern of p samples Y_i was extracted, but from the prestimulus period of the i th ST. We assessed the change in the point distribution due to the stimulus arrival, using the **R** statistic (see Appendix C) for comparing the sample $\{X_i\}_{i=1:120}$ with the sample $\{Y_i\}_{i=1:120}$. The procedure was repeated using the normalized patterns $\{X_i^*, Y_i^*\}_{i=1:120}$. The results, expressing the probability that the two point sets were coming from the same distribution, were tabulated in Table 1 and clearly favored the 'phase' encoding hypothesis.

The effect of stimulus arrival on the phase of individual STs is further visualized in Fig. 4a, where the planed MSTs of the previous two samples $\{Y_i^*\}_{i=1:120}$ and $\{X_i^*\}_{i=1:120}$ can be compared. In the prestimulus interval the phase differences among STs are random, resulting in an unorganized, uniformly distributed MST (left column). In the poststimulus interval the STs are synchronized, producing a structured MST of significantly smaller length. *These images strongly support the well-known hypothesis that stimulus has a phase resetting effect on the neuron cell activity* (Liu and Ioannides, 1995; Zouridakis et al., 1997b). Clearly, the planed MSTs of the non-normalized patterns $\{Y_i\}_{i=1:120}$ and $\{X_i\}_{i=1:120}$, shown in Fig. 4b, indicate that the stimulus arrival has no significant effect on the amplitude of STs thus supporting the phase resetting hypothesis.

3.4. Selective averages based on MST-ordering

Evaluation of the MST-ordering scheme, as a *dominant-mode seeking clustering* procedure, was performed for the unsupervised version of the algorithm and based on conventional SNR measurements (Raz et al., 1988). This evaluation further suggested an internal heuristic that can fully automate the introduced procedure so as to realize *selective averaging* of STs.

It has been shown (Laskaris et al., 1997) that given a sample $\{X_i\}_{i=1:k}$, an index of the *noise power* (np) can be estimated using the $(k \times k)$ *interpoint distance matrix* **D**:

$$np = \frac{1}{2pk(k-1)} \sum_{ij} D_{ij} \quad (3)$$

while an index of the *signal power* (sp) can easily be deduced from the average of the sample:

$$sp = \frac{1}{p} \|\bar{X}\|^2 - \frac{1}{k} np, \quad \bar{X} = \frac{1}{k} \sum_i X_i \quad (4)$$

Using these indices, estimates of SNR regarding the certain sample and the corresponding average waveform can be computed as follows:

$$SNR_{\text{sample}}(\{X_i\}_{i=1:k}) = sp/np, \quad SNR_{\text{ave}}(\{X_i\}_{i=1:k}) = k(sp/np) \quad (5)$$

In the sequel, the above two estimates are utilized in order to quantify the signal content at different levels in the ranked list of STs.

Based on the normalized patterns and rooting at the point of maximum local point density **PD**(X_i^*), the MST-ordering provides the ranked list [i], $i = 1, 2, \dots, N$ of the STs. This ordering list is then applied to the original (non-normalized) patterns X_i and the ST sample is partitioned hierarchically into subsets containing 10 STs each: $\{X_i\}_{i=j:[j+9]}$, $j = 1, 11, \dots, N - 9$. In other words, the first subset includes the STs at the top of the list, i.e. the ones having representative points X_i^* very close in \mathbf{R}^p to the root-point $X_{[1]}^*$, while the last subset includes the STs from the bottom of the list, i.e. the ones having representative points far away from the root-point. The SNR_{sample} estimate has been computed ($k = 10$), for all these subsets for the case of 120 STs

Table 1

Multivariate Wald-Wolfowitz test quantifying the stimulus-induced change in the distribution of STs, regarding their amplitude (X_i, Y_i) and phase (X_i^*, Y_i^*) representation, from the left (VS_l) and right (VS_r) virtual signal and for the three subjects (A.F., A.O., V.P.)^a

W (P value)	VS _l { X_i, Y_i } _{$i=1:120$}	VS _l { X_i^*, Y_i^* } _{$i=1:120$}	VS _r { X_i, Y_i } _{$i=1:120$}	VS _r { X_i^*, Y_i^* } _{$i=1:120$}
A.F.	-0.7 (0.24)	-3.1 (0.97×10^{-3})	-0.9 (0.18)	-4.5 (0.34×10^{-5})
A.O.	-0.5 (0.31)	-4.2 (0.13×10^{-4})	-1.6 (0.06)	-3.6 (0.16×10^{-3})
V.P.	-1.5 (0.07)	-2.7 (0.35×10^{-2})	-1.1 (0.14)	-2.1 (0.01)

^a **W** is the standardized **R** statistic and *P* value expresses the corresponding probability that the couple of point sets is coming from the same multivariate distribution.

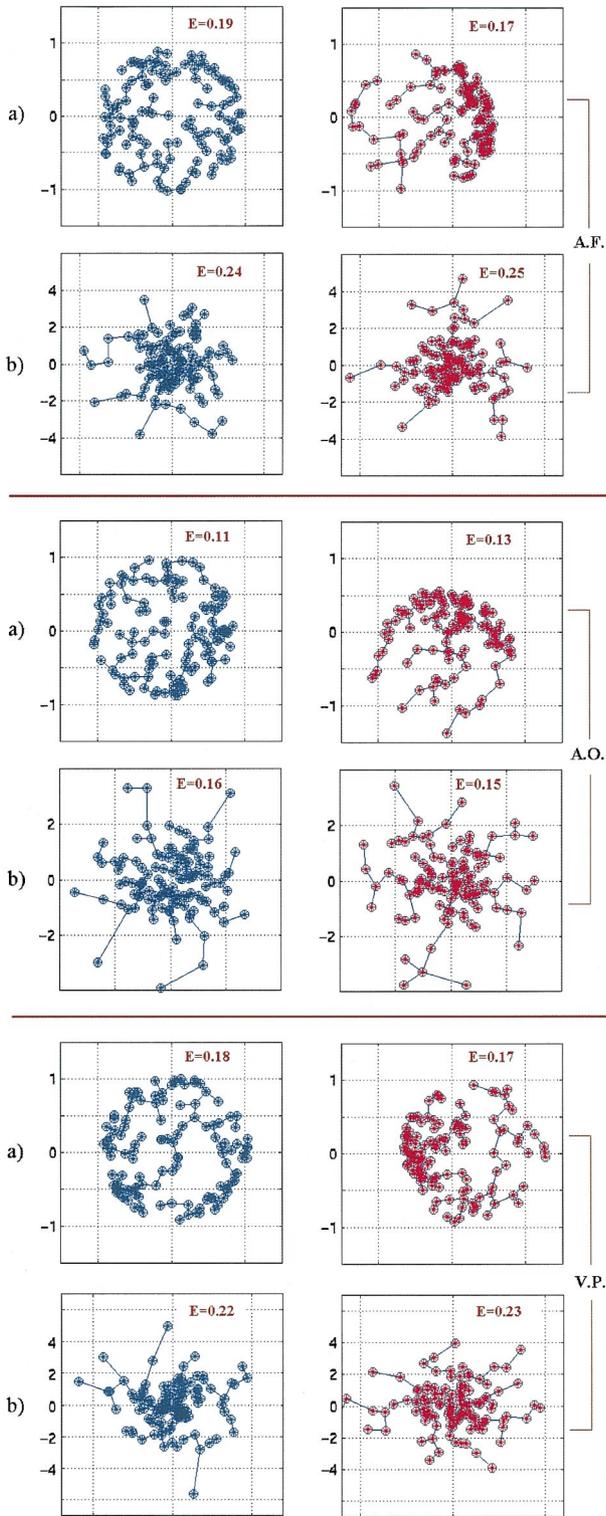


Fig. 4. Planed MSTs of patterns extracted from 120 VS₁-based STs of binaural stimulation, recorded from the three subjects. Comparisons between the distribution of STs in the prestimulus interval with their distribution in the poststimulus interval can be done by contrasting the left MST images with the right ones. Comparisons between ‘amplitude’ and ‘phase’ encoding hypotheses of the auditory response can be done by contrasting the a)-panels with the b)-ones. Similar results were obtained for the case of VS₂-based MST images.

coming from binaural stimulation and is plotted in Fig. 5a. In the same figure the SNR level as it is computed for the overall sample, $SNR_{\text{sample}}(\{X_i\}_{i=1:120})$, is indicated by the green horizontal line. A striking enhancement in SNR, induced by the MST-ordering, is evident. In Fig. 5c the

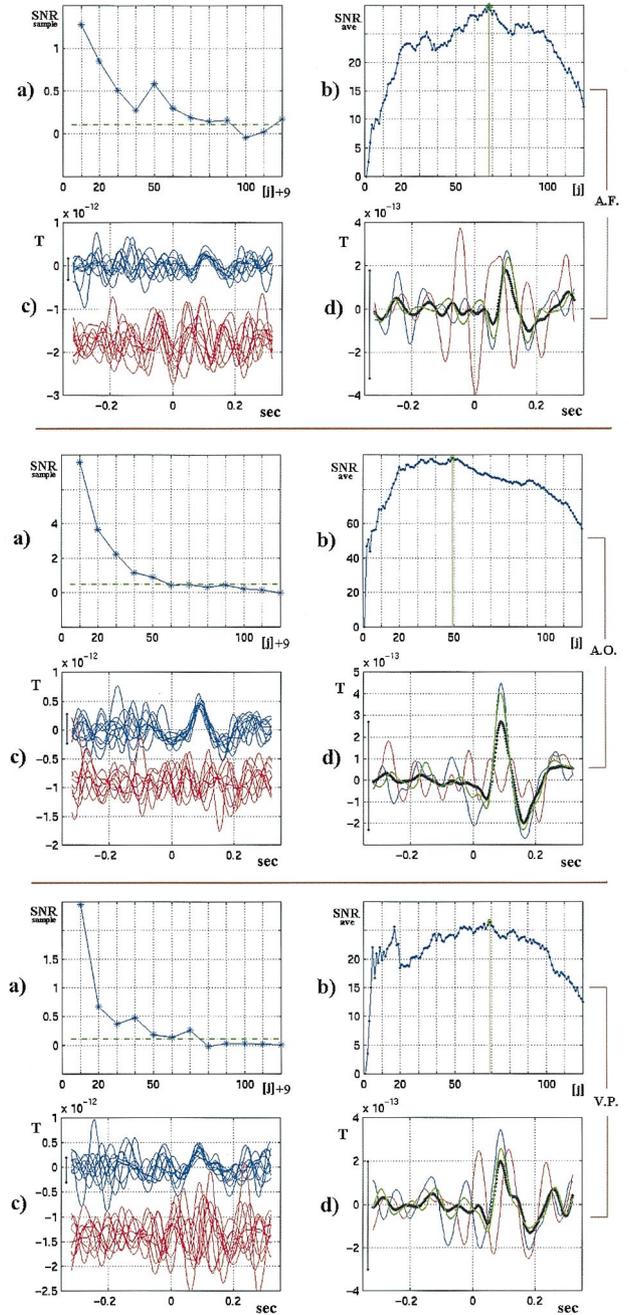


Fig. 5. SNR enhancement via the MST-ordering procedure applied to 120 VS₁-based STs of binaural stimulation, from the three subjects. (a) The sample-SNR as a function of the rank; [j] increases from the top to the bottom of the ordered list. (b) The SNR of the waveform produced by selectively averaging up to [j]-ranked ST, as a function of the rank. (c) Subsets of 10 STs cited at the top (blue) and the bottom (red) of the ordered list. (d) Averages of the 10 lowest-rank STs (blue), 10 highest-rank STs (red), J_{max} optimally selected trials of lowest-rank (light green) and all STs (black).

two subsets of STs from the top ($j = 1$) and the bottom ($j = 111$) of the ranked list have been displayed in blue and red correspondingly, showing that *the synchronization among STs, in the poststimulus time range, is what actually contributes to the above mentioned SNR enhancement*. In Fig. 5d the averages of these subsets are displayed using the corresponding colors.

The progressive deterioration of SNR, while moving towards the bottom of the STs ranked list, motivated the idea for an optimal estimate of the number of lowest-rank STs that should be included in a *selective-averaging scheme*. For the sequence of subsets $\{X_i\}_{i=[1]:[j]}$, $j = 2, 3, \dots, N$, the SNR of the corresponding average waveforms is estimated:

$$\text{SNR}_{\text{ave}}(\{X_i\}_{i=[1]:[j]}) = j * \text{SNR}_{\text{sample}}(\{X_i\}_{i=[1]:[j]}) \quad (6)$$

and the index j_{max} that maximizes the SNR_{ave} measure, provides the optimal rank for the selective average

$$\bar{X}_{\text{sel}} = \frac{1}{j_{\text{max}}} \sum_{i=[1]}^{[j_{\text{max}}]} X_i \quad (7)$$

The function of this algorithmic strategy is illustrated in Fig. 5b, where the SNR_{ave} measure is plotted as a function of rank $[j]$ and its maximum value has been denoted using a light-green circle. *The comparison of this value with the one corresponding to the ensemble average ($[j] = 120$) reveals a significant improvement*. The waveform of this selective average is depicted in Fig. 5d, drawn in light green, while the waveform of the ensemble average is drawn with a black-dotted line.

The introduced strategy, of MST-ordering followed by optimal selective averaging, will be denoted in the sequel

as *MST-SA scheme*. This scheme can be applied repeatedly to the input sample with a termination criterion based on the SNR level of the obtained average waveforms, which is a directly interpretable performance index. A detailed demonstration of this idea is given in the next section.

The validity of the hierarchy in the ST sample, established via the MST-ordering, is further demonstrated in Fig. 6, where in the left panel the two *MFT-based activation curves* from the left auditory cortex ROI, corresponding to the averages of the ST subsets at the top and the at the bottom of the ordered list, have been plotted (in blue and red correspondingly), along with the *activation curve* corresponding to the ensemble average (in black). In the right panel, the current source density distribution is visualized at the poststimulus time instant of maximum current intensity, for the average of the subset at the top of the ordered list (left) and for the ensemble average (right). Notice the difference in the maximum value of current intensity, which is depicted beneath each MRI slice.

3.5. Subtractive clustering of STs based on MST-SA scheme

The *recursive* implementation of MST-SA scheme is employed to resolve a simple classification task. The input sample contains 200 VS_r-based STs conveying the responses to a shuffled in time sequence of monaurally applied stimuli: 100 tones to each ear. The average M100 response to contralateral stimulation is known to appear earlier than the average M100 response to ipsilateral stimulation (Makela et al., 1994). In the sequel, this information is not utilized, since the aim is to justify the efficiency of the algorithm in performing unsupervised classification.

After defining the peak latency in the overall average, the

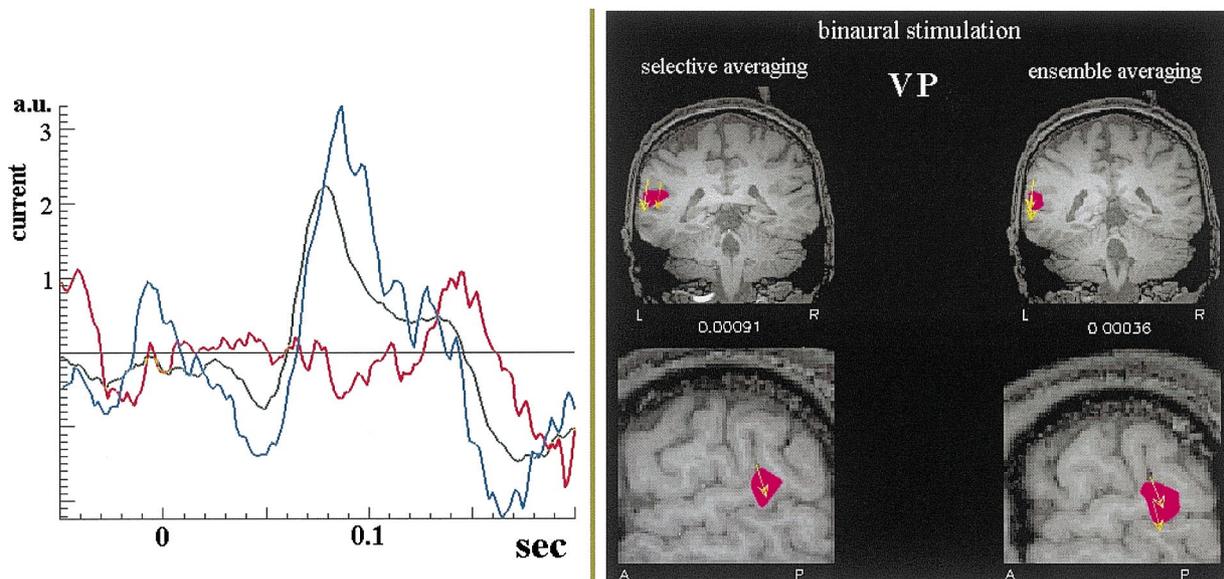


Fig. 6. (Left panel) MFT-based activation curves from a spherical ROI (1-cm radius) defined at the left auditory cortex. MFT solutions were obtained for the multichannel signal of the ensemble average (black), the average of the 10 lowest-rank STs (blue) and the average of the 10 highest-rank STs (red); the ordering of STs had been performed using the left virtual signal. (Right panel) Current source density distribution, at the poststimulus time instant of maximum current, estimated from the selective average of the 10 lowest-rank STs (left) and from the ensemble average (right).

STs are embedded as points $\{X_i^*\}_{i=1:200}$ in \mathbf{R}^p , $p = 101$. The MST is constructed and planed as shown in Fig. 7a. The overall algorithmic procedure works blindly on the unlabeled X_i^* s. However, to enable an easy justification, the label of each ST, i.e. the stimulus type, is drawn in a different color: blue(red) denotes the STs from contralateral(ipsilateral) stimuli. The local point density $\mathbf{PD}(X_i^*)$ is computed and depicted (after normalization of its maximum value to one) in the Fig. 7a, using an appended axis. This 3-dimen-

sional picture clearly shows that the ambiguity between the two classes is reflected as ‘unimodality’ in the local point density.

The above-mentioned ambiguity is due to the existence of STs with low SNR, having representative points that either populate the borders between the two regions in \mathbf{R}^p dedicated to each class, or are cited at the extremes of the point configuration. A sense of the actual situation can be obtained with the inspection of the planed ST waveforms, shown in Fig. 7b. Notice, however, that a crude classification into waveforms of early and late M100 peak can be easily performed by drawing a separating line on the plane (the vertical green lines in all the waveforms denote the latency of M100 peak in the average waveform of the contralateral responses).

To alleviate the inter-class ambiguity, a simple procedure is adopted as a signal-content enhancement preprocessing step. The STs corresponding to a \mathbf{PD} -measure below a certain level (\mathbf{PD}_0) are removed from the sample and the procedure is repeated from the MST-construction step. This preprocessing step will be fully justified in the next section, where a method for the adaptive definition of the *critical level* \mathbf{PD}_0 is also proposed. In our example, the 50% of the STs with the lowest \mathbf{PD} is removed. The new MST, of the remaining points $\{X_{i'}^*\}_{i' \in C_i}$, is planed as shown in Fig. 7c. It is obvious, from the ‘bimodality’ in the \mathbf{PD} , that the clustering task has been simplified considerably.

The MST-SA scheme has been applied after rooting at the point of maximum \mathbf{PD} , which is coming from the blue subtree. The sequence of the SNR measures $\text{SNR}_{\text{ave}}(\{X_{i'}^*\}_{i'=[1]:[j]}), j = 2, \dots, 100$ is depicted in the left panel of Fig. 7d (blue line). The rank $[j_{\text{max}}]$ maximizing this sequence is 39. The corresponding optimal average waveform, depicted in the right panel (dotted-blue line), compares favorably with the ensemble average from the whole set of STs coming from contralateral stimulation (thin-blue line). The local point density of the points corresponding to the selected 39 STs is set to zero, i.e. $\mathbf{PD}(X_{i'}^*) = 0, i' = [1]:[j_{\text{max}}]$, and the second important cluster is delineated, via the MST-SA scheme, after rooting at the maximum of the modified \mathbf{PD} . The new sequence of SNR measures, the corresponding selective average and the ensemble average from the whole set of STs coming from ipsilateral stimulation are presented in Fig. 7d, drawn in red. The third iteration results in an average waveform of very low SNR_{ave} index; approximately 1/21 of the same index estimated for the original input sample (of 200 STs). The overall subtractive-clustering scheme was applied also to the corresponding sample of VS_I -based STs, showing similar performance.

The effectiveness of the suggested algorithmic procedure, its computational simplicity (since only simple manipulations of the distance matrix elements are involved) and the inherent quality assurance strategy (optimality in SNR sense is guaranteed) make this technique very promising for extracting *prototypical responses* from the ST samples.

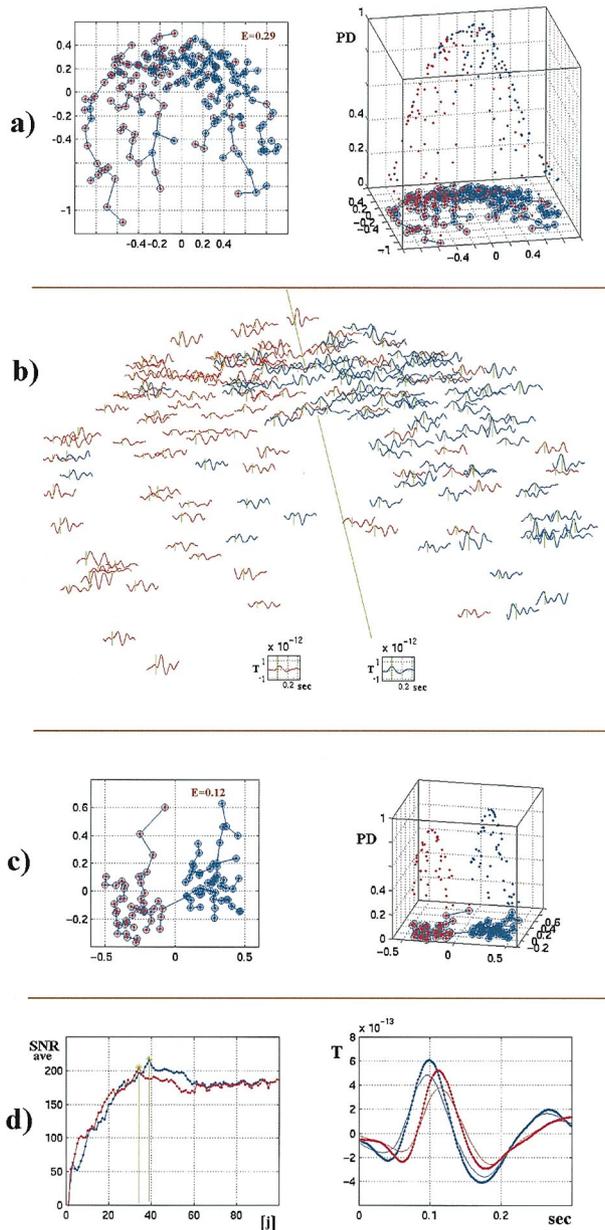


Fig. 7. Algorithmic steps of the MST-SA based subtractive clustering technique, applied to a sample of 200 VS_I -defined STs: 100 from contralateral and 100 from ipsilateral stimulation. The (average) M100 contralateral response is known to appear earlier than the ipsilateral one. Blue (red) color denotes points/ST-waveforms/ SNR_{ave} -curves/selective averages corresponding to the subset of 100 STs conveying the contralateral (ipsilateral) response. For a detailed description, see Section 3.5.

Note for example how the prototypes, in our example of Fig. 7d, vividly demonstrate the ipsilateral versus contralateral delay of the M100 peak.

3.6. SNR-Monitoring using local point density measures

The use of local point density measure **PD**, as an index of instantaneous SNR, is demonstrated in Fig. 8a, for a sample of 120 STs coming from the binaural stimulation of the subject A.O. From each ST, a couple of p -length patterns $\{X_i, Y_i\}$ was extracted; X_i was defined from the poststimulus interval around the peak-latency in the averaged M100 response ($p = 101$), while Y_i came from the prestimulus interval.

The MST of (normalized patterns) $\{X_i^*\}_{i=1:120}$ was constructed and its length l_{120} provided the radius r_0 for the estimation of local concentration in both $\{X_i^*\}_{i=1:120}$ and $\{Y_i^*\}_{i=1:120}$ point- samples (see Appendix A):

$$PD(X_i^*) = \frac{1}{(2\pi)^{p/2} r_0^p N} \sum_{j=1}^N \exp\left[-\frac{\|X_i^* - X_j^*\|^2}{2r_0^2}\right], \quad PD(Y_i^*)$$

$$= \frac{1}{(2\pi)^{p/2} r_0^p N} \sum_{j=1}^N \exp\left[-\frac{\|Y_i^* - Y_j^*\|^2}{2r_0^2}\right], \quad N = 120, \quad p = 101 \quad (8)$$

The sequences of both $PD(X_i^*)$ and $PD(Y_i^*)$ measures have been plotted (in red and blue, respectively), after scaling with the maximum value of $PD(X_i^*)$, as a function of i , i.e. the time identity of each ST. In this way this geometrical point-set descriptor, computed in the p -dimensional space, has been ‘projected’ on the trial-to-trial dimension. Left (right) panel of Fig. 8 corresponds to the signal from the left (right) hemisphere. It is easy to observe: *the great difference in the mean value between the two sequences, the great fluctuations in the poststimulus defined PD and the considerable number of poststimulus points having a PD-measure as low as the PD of prestimulus points*. The same features are present in the curves, obtained from the other subjects’ data (not shown).

The common radius, in the computation of both PD-estimates (Eq. (8)), acts as a normalization factor, making the $PD(Y_i^*)$ values serve as a ‘baseline distribution’ for the values of local point density of a randomly distributed point-sample. Whenever the sequence of $PD(Y_i^*)$ is station-

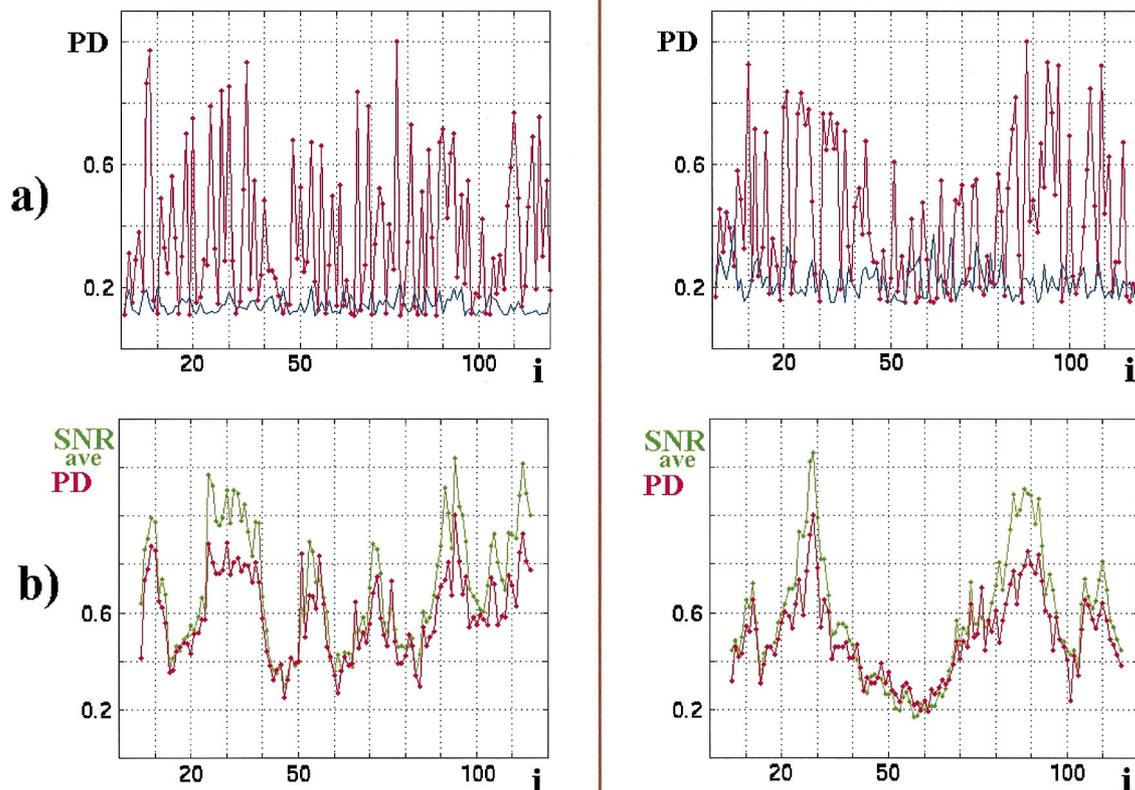


Fig. 8. Evolution of local point density (**PD**) along the trial-to-trial dimension, for a sample of 120 STs from binaural stimulation. Left (right) panel shows the results obtained using the virtual signal from the left (right) hemisphere. (a) **PD** as a function of the ST time-identity. For each ST two **PD**-estimates were evaluated; one was based on normalized patterns extracted from the prestimulus and the other from the poststimulus time interval. The corresponding **PD**-sequences are depicted as blue and red line, respectively. (b) Sequence of quasi-instantaneous SNR measurements (green), based on moving averages and ‘smoothed’ version of the sequence of **PD**-estimates (red) based on the same moving averages.

ary over the recording session, its mean value can provide the critical level \mathbf{PD}_0 , that above which the local point concentration indicates stimulus induced self-organization. Formally, a *sigmoid-type transformation*, using the \mathbf{PD}_0 as threshold, can be employed. Note that under the adopted signal representation, this critical level practically quantifies the aggregate synchronization –in time- among the STs in the prestimulus range.

In order to provide an indication that the \mathbf{PD} measure can serve as an estimator of the instantaneous-SNR, we show its ‘equivalence’ –in performance- with the conventional SNR_{ave} estimator, when the latter acts in a moving-average mode along the trial-to-trial dimension (Laskaris et al., 1997; Jung et al., 1999). A window, of depth $w = 11$, slides along the stack of the ST patterns $\{X_i\}_{i=1:120}$ enclosing at the i th step the subset $\{X_j\}$, $j = i - (w - 1)/2 : i + (w - 1)/2$ and producing as output the SNR measurement $\text{SNR}_{\text{ave}}^i = \text{SNR}_{\text{ave}}(\{X_j\})$ (Eqs. 3–5 with $k = w$), based on the corresponding local average waveform

$$X_{\text{ave}}^i = \frac{1}{w} \sum_j X_j \quad (9)$$

This $\text{SNR}_{\text{ave}}^i$ value, as a function of i , is drawn (green curve) in Fig. 8b. Since this index actually measures the ‘local average’ SNR behavior, the corresponding curve can not be compared directly with the curve of $\mathbf{PD}(X_i^*)$ values (Fig. 8a). It is expected to be a ‘low pass’ version of the successive \mathbf{PD} measurements, as it can easily be verified by inspection. In Fig. 8b we have also drawn, after normalizing the maximum value to one, the curve of \mathbf{PD} measurements over the sample of the above waveforms $\{X_{\text{ave}}^i\}_{i=6:115}$, produced by the moving average filter. The comparison of the two curves clearly shows the ‘duality’ of the two SNR-monitoring techniques, whenever the \mathbf{PD} -based one is applied in a quasi-instantaneous manner.

Tracking the SNR via the proposed technique can prove to be a very informative aspect regarding the course of evoked response. Fig. 8 serves as a characteristic example of this. The included graphs (and the omitted ones corresponding to the other two subjects), from a time-series analysis point of view, indicate: *the superposition of both stochastic and deterministic components, independence between the two hemispheres regarding the stochastic components of the curves and (weak) coupling between them regarding the deterministic components*. The full understanding of these graphs necessitates the simultaneous study of signals from other brain areas and is beyond the scope of this work.

Generally speaking, changes in the \mathbf{PD} index, under verified ‘stationary’ ongoing activity, can be interpreted as changes in the signal content, and their monitoring can indicate/suggest (the existence of) possible deterministic laws governing the response generation. Compensation for \mathbf{PD} changes, irrelevant to the response generation mechanism (e.g. due to changes in the level of arousal), could be sought

by incorporating information from more channels (e.g. so as to adaptively adjust the critical level \mathbf{PD}_0 in a *sigmoid type transformation* of the \mathbf{PD} -measurements).

4. Discussion

We have introduced an EDA framework utilizing well established graph-theoretic concepts to capture the structure of the ST sample. The novelty of the approach is the description of the input sample in the form of self-organized ordered lists, offering a powerful and theoretically well founded way for the quantification and understanding of the trial-to-trial variation.

The validation of the established hierarchies was, intentionally, based on data from well-understood magnetoencephalographic experimental paradigms. The problem of biologically induced artifacts was not addressed in this validation, because of the good noise elimination properties of the employed virtual signal. While the benefits from the application of this internal classification need to be explored further (with more complicated tasks), its simplicity, optimality and ability to provide insights (through the attached visualization scheme) make it an attractive ‘first choice’ for understanding the input sample. Moreover, the methodologies are expected to share the robustness to artifacts which is a characteristic inherent to all the ordering based processing schemes (Bezerianos et al., 1995; Laskaris et al., 1996a).

Although the emphasis had been placed on the methodology, its application to real data produced intriguing results deserving further exploration. First, a clear support for the hypothesis that the stimulus induces a phase resetting of the signal in the 3–20 Hz band. In an earlier study specializing on the stimulus evoked 40 Hz activity, it was shown that the weak gamma band activity surviving after averaging thousands of trials (Pantev et al., 1991; Ribary et al., 1991) is generated by the resetting in a few percentage of trials and sharpening of an ongoing process that occurs intermittently and spontaneously in single trials well before and well after the stimulus onset (Liu and Ioannides, 1995). Second, the inhomogeneity in the ST samples and especially the existence of STs distinct from the majority of them. Third, the existence of trial-to-trial systematic variation as trials progress which underlies the overall structure of ST-samples. The use of dual time, one within trials and the other across the succession of trials, seems a promising direction.

During the application of the proposed methodology the STs were represented by time-locked patterns. This representation is susceptible to *latency-jitter* which may introduce its own signature into the MST structure. This would be readily recognized during the navigation in the ST sample and (if necessary) could be overcome by incorporating a peak-detection algorithm to identify independently the peak latency for each ST prior the pattern extraction

step. Effects due to latency-jitter were not significant in the specific examples we have employed in our study.

The presented EDA framework was introduced using a ‘mono-channel’ signal, the virtual signal. Since, the framework aims to the explore the *trial-to-trial dimension*, it can easily be modified so as to handle multi-channel signals. The most direct way, fully compatible with the previous presentation, is the extraction of different patterns from each channel and the inclusion of them in a composite pattern. In this way, each ST will be represented as a point in a multidimensional space which is the Cartesian product of the individual spaces corresponding to the mono-channel pattern extraction. The most conventional way is the incorporation of a dimensionality reduction technique, as a preprocessing step, that either works blindly on the data (e.g. PCA/ICA) or is guided by neuroanatomical considerations (e.g. linear combinations of channels from specific brain areas). In this way a mono-dimensional signal can be identified as the carrier of evoked response (Lukka et al., 2000) and the EDA framework will be applied on it.

Finally, the introduction of self-organized ordering of STs constitutes a promising way for exploring nonlinear interactions among channels/brain areas. The ordering of a ST sample can be performed for two (or more) different channels independently, and since ranks reflect the signal content, similarity in the ranked lists can be interpreted either as common signal content in or dependency among the two channels.

Appendix A

Given a sample $\{X_i\}_{i=1:N}$ of points in \mathbf{R}^p , the *local point density* \mathbf{PD} at each sample point X_i can be estimated by counting the number of points included in a hypersphere centered at X_i :

$$V_i = \#\{Y : \|X_i - Y\| \leq r, Y \in \{X\}_{i=1:N}\}$$

and normalizing with the sample size, i.e. $\mathbf{PD}(X_i) = V_i/N$.

Recently (Chaudhury et al., 1996), the previous \mathbf{PD} estimation was related with the construction of MST for the given sample. It was shown that the optimum radius r_0 for this estimation is related with the length l_N of the MST, i.e. the sum of its *edge weights*, with the equation:

$$r_0 = \left(\frac{l_N}{N}\right)^{1/p}$$

The smooth version of this \mathbf{PD} estimator is cast in a form of *potential functions* with r_0 shaping the influence of each point on the rest (Parzen, 1962; Cacoullos, 1966):

$$\mathbf{PD}(X_i) = \frac{1}{(2\pi)^{p/2} r_0^p N} \sum_{j=1}^N \exp\left[-\frac{\|X_i - X_j\|^2}{2r_0^2}\right]$$

Appendix B

The MST-based mapping of a sample of points $\{X_i\}_{i=1:N} \in \mathbf{R}^p$ to a sample of points on a plane $\{\chi_i\}_{i=1:N}$ is summarized within the following steps:

step 1: the $(N \times N)$ *interpoint distance matrix* is computed: $\mathbf{D} = \text{diag}(\mathbf{A})\mathbf{E} + \mathbf{E}\text{diag}(\mathbf{A}) - 2\mathbf{A}$, where

$$D_{kl} = \|X_k - X_l\|^2, \mathbf{A} = \mathbf{X}\mathbf{X}^T, \mathbf{E}_{(N \times N)} = \begin{pmatrix} 11\dots 1 \\ \vdots \\ 11\dots 1 \end{pmatrix}, \mathbf{X}_{(N \times p)} = [X_1 \dots X_N]$$

step 2: The MST is constructed. A classical algorithm for this construction (Kruskal, 1956), utilizes the set of *interpoint distances* D_{kl} as the search-space and proceeds in an iterative manner using two steps: (i) selection of the shortest edge, which does not form any loops with those edges already chosen, and (ii) removal of the selected edge from the search-space.

step 3: The MST is rooted at a selected point and MST-ordering is performed.

step 4: Having mapped the root point $X_{[1]}$ to $\chi_{[1]} = (0, 0)$ and $X_{[2]}$ to $\chi_{[2]} = (0, e(X_{[1]}))$, the mapping rule for the remaining points $X_{[i]}$, $i > 2$ in the ordered list, takes the form:

$$\text{if } P(X_{[i]}) \mapsto (\chi_p, \psi_p), \text{ then } X_{[i]} \mapsto \chi_{[i]} = (\chi, \psi) :$$

$$\chi = \frac{d13}{d12}(\cos\varphi\chi_p - \sin\varphi\psi_p), \psi = \frac{d13}{d12}(\sin\varphi\chi_p + \cos\varphi\psi_p)$$

where

$$\cos\varphi = \frac{d13^2 + d12^2 - d23^2}{2d13d12}, \quad d13 = \|X_{[1]} - X_{[i]}\|, \quad d12 = \|X_{[1]} - P(X_{[i]})\|, \quad d23 = \|P(X_{[i]}) - X_{[i]}\| = e(X_{[i]})$$

Appendix C

Given two multidimensional point samples $\{X_i\}_{i=1:N}$ and $\{Y_i\}_{i=1:N}$, the hypothesis H_0 to be tested is whether they are coming from the same multivariate distribution. At first, the sample identity of each point is not encountered and the MST of the overall sample is constructed. Then, based on the sample identities of the points, a test statistic R is computed. R is the total number of *runs*. A *run* is defined as a consecutive sequence of identical sample identities. Rejection of H_0 is for small values of R . The null distribution of the test statistic has been derived, based on combinatorics. It has been shown that the quantity

$$W = \frac{R - E[R]}{\sqrt{\text{Var}[R]}}$$

approaches (asymptotically) the standard normal distribution, where the $E[R]$ and $\text{Var}[R]$ can be computed based on the size of the two samples (Friedman and Rafsky, 1979).

This enables the computation of the *significance level* (and *P value*) for the acceptance of the hypothesis H_0 .

References

- Astola J, Haavisto P, Neuvo Y. Vector median filters. *Proc IEEE* 1990;78(4):678–689.
- Barnett V. The ordering of multivariate data. *J R Statist Soc A* 1976;139:318–354.
- Bezerianos A, Laskaris N, Fotopoulos S, Papathanasopoulos P. Data dependent weighted averages for recording of evoked potential signals. *Electroencephalogr Clin Neurophysiol* 1995;96:468–471.
- Cacoullos T. Estimation of a multivariate density. *Ann Inst Math Statist* 1966;18:179–189.
- Cansino S, Tellez-Alanis B. ERPs elicited by a cognitive incongruity paradigm: a semantic memory study. *Neuroreport* 2000;11(5):977–981.
- Chaudhury D, Chaudhury BB, Murthy CA. A data driven procedure for density estimation with some applications. *Pattern Recogn* 1996;10:1719–1736.
- Chowdhury N, Murthy CA. Minimal spanning tree based clustering technique: relationship with bayes classifier. *Pattern Recogn* 1997;11:1919–1929.
- Friedman J, Rafsky L. Multivariate Generalizations of the Wald-Wolfowith and Smirnov two-sample tests. *Ann Statist* 1979;7:697–717.
- Friedman J, Rafsky L. Graphics for the multivariate two-sample problem. *J Am Statist Assoc* 1981;76:277–287.
- Gath I, Bar-On EH, Lehmann D. Automatic classification of visual evoked responses. *Comput Methods Progr Biomed* 1985;20:17–22.
- Geva A. Feature extraction and state identification in biomedical signals using hierarchical fuzzy clustering. *Med Biol Eng Comput* 1998;36:608–614.
- Geva A, Pratt H. Unsupervised clustering of evoked potentials by waveform. *Med Biol Eng Comput* 1994;32:543–550.
- Hardie R, Arce G. Ranking in R^p and its use in multivariate image estimation. *IEEE Trans Circuits Syst Video Technol* 1991;1(2):197–209.
- Hartigan J. Clustering algorithms, New York: Wiley, 1975. pp. 201–205.
- Ioannides A. Continuous probabilistic solutions to the biomagnetic inverse problem. *Inverse Problem* 1990;6:523–542.
- Ioannides A, Hasson R, Miseldine G. Model-dependent noise elimination and distributed source solutions for the biomagnetic inverse problem. *Proc SPIE* 1990;1351:471–481.
- Jain A, Dubes R. Algorithms for clustering data, New Jersey: Prentice Hall, 1988. pp. 55–141.
- Jung TP, Makeig S, Westerfield M, Townsend J, Courchesne E, Sejnowski T. Analysing and visualizing single-trial event-related potentials. *Adv Neural Inform Process Syst* 1999;11:118–124.
- Kruskal J. On the shortest spanning subtree of a graph and the traveling salesman problem. *Proc Am Math Soc* 1956;7:48–50.
- Kruskal J. Nonmetric multidimensional scaling: a numerical method. *Psychometrika* 1964;29(2):115–129.
- Lange D, Siegelmann H, Pratt H, Gideon I. Overcoming selective ensemble averaging: unsupervised identification of event-related brain potentials. *IEEE Trans Biomed Eng* 2000;47:822–826.
- Laskaris N, Bezerianos A, Fotopoulos S. Unsupervised artifact rejection in EP recordings by means of a novel nonlinear technique. *Appl Signal Process* 1996a;3:150–154.
- Laskaris N, Bezerianos A, Fotopoulos S, Papathanasopoulos P. A genetic algorithm for the multidimensional filtering process of visual evoked potential signals. 18th IEEE EMBS Conference, Amsterdam, Netherlands, 1996, article no. 5.8.3-5 (CD-ROM).
- Laskaris N, Fotopoulos S, Papathanasopoulos P, Bezerianos A. Robust moving averages, with Hopfield Neural Network implementation, for the monitoring of evoked potential signals. *Electroencephalogr Clin Neurophysiol* 1997;104:151–156.
- Lee R, Slagle J, Blum H. A triangulation method for the sequential mapping of points from N-space to two-space. *IEEE Trans Comput* 1977;C26:288–292.
- Liu L, Ioannides A. Single epoch analysis of MEG signals. In: Baumgartner C, Deeke L, Stroink G, Williamson SJ, editors. *Biomagnetism: fundamental research and clinical applications*, Amsterdam: Elsevier Science, 1995. pp. 439–444.
- Liu L, Ioannides A. A correlation study of averaged and single trials MEG signals: the average describes multiple histories each in a different set of single trials. *Brain Topogr* 1996;8(4):385–396.
- Liu L, Ioannides A, Muller-Gartner H. Bi-hemispheric study of single trial MEG signals of the human auditory cortex. *Electroencephalogr Clin Neurophysiol* 1998a;106:64–78.
- Liu L, Ioannides A, Taylor J. Observation of quantization effects in human auditory cortex. *Neuroreport* 1998b;9:2679–2690.
- Lukka T, Schoner B, Marantz A. Phoneme discrimination from MEG data. *Neurocomputing* 2000;31:153–165.
- Makela J, Hamalainen M, Hari R, McEvoy L. Whole-head mapping of middle latency auditory evoked magnetic fields. *Electroencephalogr Clin Neurophysiol* 1994;92:414–491.
- Pantev C, Makeig S, Hoke M, Galambos R, Hampson S, Gallen C. Human auditory evoked gamma-band magnetic fields. *Proc Natl Acad Sci USA* 1991;88:8996–9000.
- Parzen E. On estimation of a probability density function and mode. *Ann Math Statist* 1962;33:1065–1076.
- Rama P, Paavilainen L, Anourova I, Alho L, Reinikainen K, Sipila S, Carlson S. Modulation of slow brain potentials by working memory load in spatial and nonspatial auditory tasks. *Neuropsychologia* 2000;38(7):913–922.
- Raz J, Turetsky B, Fein G. Confidence intervals for signal-to-noise ratio when a signal embedded in noise is observed over repeated trials. *IEEE Trans Biomed Eng* 1988;36:646–649.
- Ribary U, Ioannides A, Singh KD, Hasson R, Bolton J, Lado F, Mogilner A, Llinas R. Magnetic field tomography (MFT) of coherent thalamo-cortical 40 Hz oscillations in humans. *Proc Natl Acad Sci USA* 1991;88:11037–11041.
- Sindoukas D, Laskaris N, Fotopoulos S. Algorithms for color image edge enhancement using potential functions. *IEEE Signal Process Lett* 1997;4:269–272.
- Yager R, Filev D. Generation of fuzzy rules by mountain clustering. *J Intell Fuzzy Syst* 1994;2:209–219.
- Wackermann J, Matousek M. From the ‘EEG age’ to a rational scale of brain electric maturation. *Electroencephalogr Clin Neurophysiol* 1998;107(6):415–421.
- Zahn C. Graph-theoretical methods for detecting and describing gestal clusters. *IEEE Trans Comput* 1977;C20:68–86.
- Zouridakis G, Boutros N, Jansen B. A fuzzy clustering approach to study the auditory P50 component in schizophrenia. *Psychiatry Res* 1997a:169–181.
- Zouridakis G, Jansen B, Boutros N. A fuzzy clustering approach to EP estimation. *IEEE Trans Biomed Eng* 1997b;44:673–680.