

**ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ
ΚΡΗΤΗΣ**

Σχολή Τεχνολογικών Εφαρμογών
Τμήμα Εφαρμοσμένης Πληροφορικής & Πολυμέσων

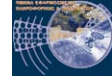
**‘Επεξεργασία MEG με BSS/ICA και
χρήση Predictive model’**

-Πτυχιακή εργασία-

“Παναγιώτης Νικολόπουλος Α.Μ.: 390”

- Ηράκλειο 2006 -





Πρόλογος

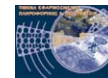
Η εργασία πραγματοποιήθηκε στο Εργαστήριο Ευφυϊών Συστημάτων του Τμήματος Εφαρμοσμένης Πληροφορικής & Πολυμέσων του Τεχνολογικού Εκπαιδευτικού Ιδρύματος Κρήτης κατά το έτος 2006, κάτω από την επίβλεψη του καθηγητή Γ.Παπαδουράκη.



Ευχαριστίες

Για την επιστημονική καθοδήγηση, την εμπιστοσύνη και τη συνεργασία τους θα ήθελα να ευχαριστήσω τους Prof. Γ.Παπαδουράκη, Ph.D. Μ.Βούρκα, Prof B.Jervis και Ph.D. S.Belal μέλη του Biopattern Research Group.

Χωρίς την συμβολή των παραπάνω ατόμων δεν θα ήταν δυνατή η υλοποίηση της εν λόγω εργασίας.



Περίληψη

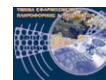
Η συγκεκριμένη εργασία πραγματοποιήθηκε με σκοπό την ανάπτυξη μιας μεθόδου εξόρυξης των Προκλητών Δυναμικών-EPs που περιέχονται στα ICs (τις Ανεξάρτητες Συνιστώσες) από κάθε πειραματική δοκιμή.

Τα EPs είναι αποκρίσεις σημάτων μικρου πλάτους, τα οποία δημιουργούνται κατά τη διάρκεια των πειραμάτων μνήμης. Τα EPs είναι γραμμικές μείξεις με μεγάλα σήματα, όπως το εγκεφαλογράφημα (EEG) και ο θόρυβος, που διακινούνται εντός του κρανίου. Ο Διαχωρισμός Πηγών στα Τυφλά (BSS) με χρήση της Αναλυσης Ανεξάρτητων Πηγών (ICA) επιτυγχάνεται σύμφωνα με τη θεωρία περί Μεγιστοποίησης της Πληροφορίας (InfoMax). Ο InfoMax χρησιμοποιείται ώστε να διαχωρίσει τις συνιστώσες των EPs από τα καταγεγραμμένα σήματα.

Η ICA είναι μια τεχνική BSS που ανακτά N ανεξάρτητες (χρονικά) πηγές $\mathbf{s} = [s_1(t), \dots, s_n(t)]^T$ από N γραμμικές μείξεις, $\mathbf{x} = [x_1(t), \dots, x_m(t)]^T$. Οι μείξεις είναι αποτέλεσμα του πολλαπλασιασμού του πίνακα των άγνωστων πηγών \mathbf{s} με ένα αγνωστο πίνακα \mathbf{A} (πίνακα μείξης), $\mathbf{x} = \mathbf{A} \cdot \mathbf{s}$. Η ICA, μη έχοντας γνώση για τις αρχικές πηγές, αναζητά ένα τετραγωνικό πίνακα \mathbf{W} (διαχωριστή/φίλτρο ή πίνακα διαχωρισμού), έτσι οι οι πηγές υπολογίζονται ως εξής: $\mathbf{s} = \mathbf{W} \cdot \mathbf{x}$.

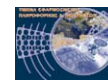
Εκμεταλευόμενοι τα συγκριτικά πλεονεκτήματα που προσφέρει η ICA, εφαρμόζεται μία μέθοδος εξόρυξης Προκλητών Δυναμικών με χρήση χρονικών και τοπογραφικών χαρακτηριστικών των ανεξάρτητων συνιστωσών.

Η μέθοδος επεξεργάζεται τεχνητά δεδομένα με ήδη γνωστές πηγές και επιστρέφει ομαδοποιημένες προβολές με σκοπό τη διατήρηση μόνο των σημαντικών συνιστωσών που περιέχουν την πληροφορία που παρουσιάζει ιατρικό ενδιαφέρον.



Περιεχόμενα

Πρόλογος	2
Ευχαριστίες	3
Περίληψη	4
Περιεχόμενα	5
1 Λεξιλόγιο	7
2 Εισαγωγή	8
3 Το θεωρητικό υπόβαθρο της εργασίας	11
3.1 Διαχωρισμός Πηγών στα Τυφλά (BSS)	11
3.1.1 Χωρική BSS	13
3.1.2 Χρονική BSS	14
3.2 Ανάλυση Ανεξάρτητων Συνιστωσών (ICA)	14
3.3 Εφαρμόζοντας την ICA σε δεδομένα EEG	21
3.4 InfoMax	23
3.5 Περιβάλλον υλοποίησης και πακέτα λογισμικού	26
4 Πειραματική διαδικασία	27
4.1 Περιγραφή Δεδομένων	27
4.1.1 Location file (Αρχείο Ηλεκτροδίων)	27
4.1.2 Τεχνητά σήματα EEG με γνωστές πηγες	28
4.2 Μεθοδολογία ανάλυσης δεδομένων	30
4.2.1 Επιλογή Συνιστωσών (ICs)	32
4.2.1.1 Μέγιστη Διακύμανση	32
4.2.1.2 Πίνακας Μείξης	32
4.2.1.3 Zero-crossings	32

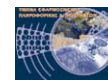


4.2.2	Επιλογή Χαρακτηριστικών για Ομαδοποίηση	32
4.2.3	Διαδικασία Ομαδοποίησης	33
4.3	Περιγραφή αποτελεσμάτων	34
4.3.1	Διαδικασία προεπεξεργασίας δεδομένων	34
4.3.2	Ανάλυση και οπτικοποίηση πριν την ICA	35
4.3.3	Οπτικοποίηση του μέσου EP	38
4.3.4	Εντοπισμός της μέγιστης διακύμανσης	40
4.3.5	Απόρριψη zero-crossings	41
4.3.6	Εξαγωγή χαρακτηριστικών για τα δύο στάδια ομαδοποίησης	42
4.3.7	1 ^η ομαδοποίηση – Χαρακτηριστικά των κορυφών	43
4.3.8	2 ^η ομαδοποίηση - Στήλες των πινάκων μείξης	45
4.3.8.1	Οπτικοποίηση τελικών υπο-ομάδων	46
4.3.8.2	Οι υπο-ομάδες που αποτελούνται από τις αρχικές πηγές	49
4.3.8.3	Παράδειγμα υπο-ομάδας παρασιτικών πηγών	53
4.4	Σχολιασμός αποτελεσμάτων	55
5	Κριτική	56
6	Επίλογος	57
7	Βιβλιογραφία	58
8	Παρατήματα	62
8.1	Δείγμα κώδικα υλοποίησης	62
8.2	Προσέγγιση διπολικών πηγών	68



1 Λεξιλόγιο

BSS	Blind Source Separation (Διαχωρισμός Πηγών στα Τυφλά)
ICA	Independent Component Analysis (Ανάλυση Ανεξάρτητων Συνιστωσών)
EP	Evoked Potential (Προκλητό Δυναμικό)
Cz	Όνομα ηλεκτροδίου
EEG	Ηλεκτροεγκεφαλογράφημα
InfoMax	Ένας αλγόριθμος υλοποίησης της ICA
IC	Independent Component (Ανεξάρτητη Συνιστώσα)
MI	Mutual Information (Κοινή πληροφορία)
ML	Maximum-Likelihood (Μέγιστη Πιθανότητα)
PDF	Probability Density Function (Συνάτηση Πυκνότητας Πιθανότητας)
FastICA	Ένας αλγόριθμος υλοποίησης της ICA
EEGLAB	Βασική βιβλιοθήκη για την υλοποίηση
ICAEPtoolbox	Βασική βιβλιοθήκη για την υλοποίηση
Trial	Μία ανεξάρτητη πειραματική δοκιμή



2 Εισαγωγή

Τα βιο-ιατρικά σήματα που προέρχονται από πολλές πηγές όπως καρδιά, εγκέφαλο και ενδοκρινές σύστημα προκαλούν μία πρόκληση για του ερευνητές που καλούνται να διαχωρίσουν τα ασθενή σήματα που προέρχονται από πολλαπλές πηγές που παρεμβάλλονται από παράσιτα και θόρυβο. Η ανάλυση αυτών των σημάτων είναι εξαιρετικά σημαντική για την έρευνα όσο και για την ιατρική διάγνωση και θεραπεία.

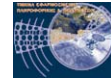
Οι εφαρμογές της ICA στα βιο-ιατρικά σήματα έχουν εξαπλωθεί πολύ γρήγορα διευρύνοντας το πεδίο έρευνας και πολλές ομάδες δραστηριοποιούνται στην αποσυνέλιξη σημάτων και στον διαχωρισμό σημάτων στα τυφλά (BSS), με σκοπό την απόκτηση νέων πληροφοριών για τον εγκέφαλο και το σώμα.

Σε αυτή τη εργασία χρησιμοποιείται η μέθοδος ICA σε δεδομένα με EPs, από καταγραφές EEG. Στόχος είναι η ανάλυση και η απομόνωση των EPs και των πηγών για κάθε πειραματική δοκιμή.

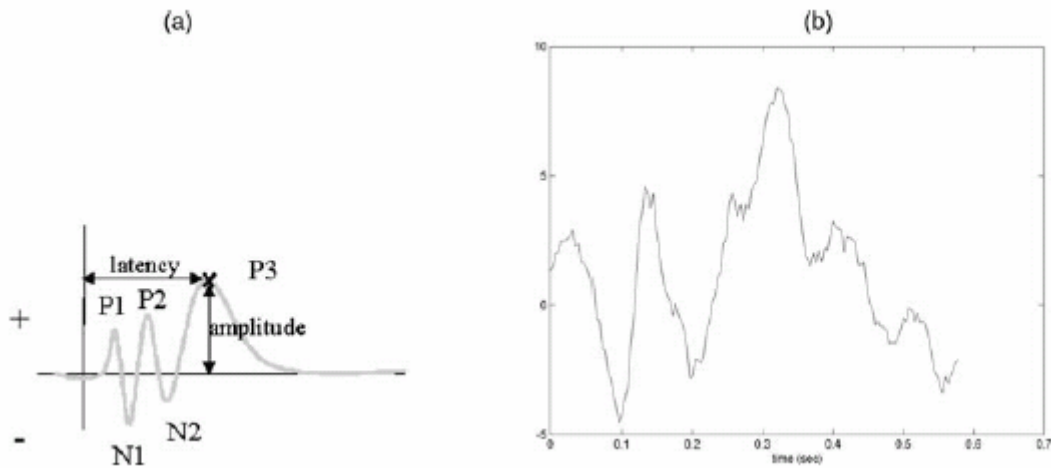
Τα EPs είναι ηλεκτρικά σήματα που παράγονται από αποκρίσεις του εγκεφάλου λόγω ερεθίσματος ή νοητικής δραστηριότητας. Τα EPs είναι πολύ χρήσιμα για την εξερεύνηση του εγκεφάλου γιατί προέρχονται από συγκεκριμένες περιοχές, παρουσιάζουν καθορισμένο σχήμα και επηρεάζονται από ήδη γνωστούς τρόπους και κάτω από συγκεκριμένες συνθήκες.

Από την άλλη τα σήματα EEG έχουν τυχαίο σχήμα, καταγράφονται από καθορισμένες θέσεις και αποτελούν το αθροισμα των επιρροών πλήθους σημάτων από διάφορες περιοχές. Τα EPs πάντα περιέχονται αναμεσά στα EEGs.

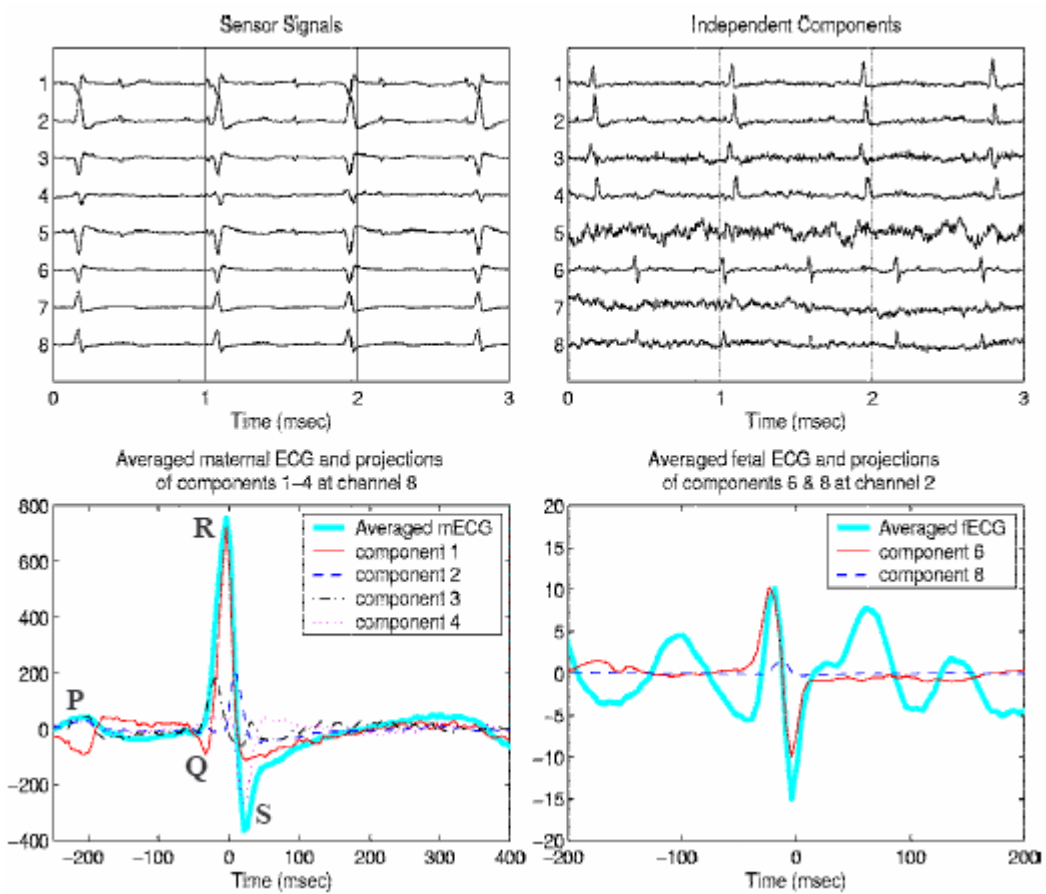
Εξαιτίας της τυχαίας φύσης των EEGs και σε αντιδιαστολή με τη σχετικά μικρή διάρκεια και χαμηλή τάση των EPs, έχει εφαρμοστεί ως ιατρική πρακτική η χρήση του μέσου όρου από πολλές πειραματικές δοκιμές για τον υπολογισμό του προκλητού δυναμικού-EP. Ο υπολογισμός του μέσου δίνει έμφαση στο EP και μειώνει την παρουσία του EEG. Τα κοινά EPs χαρακτηρίζονται από δύο μόνο παράμετρους, την



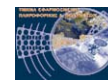
τάση της κορυφής του και την χρονική στιγμή που αυτή συμβαίνει (βλ. Σχ.1). Αυτός ο χαρακτηρισμός όμως είναι σχετικός και ανακριβής.



Σχ.1 α) Ένδειξη των χαρακτηριστικών ενός EP. β) Παράδειγμα ενός προκλητού δυναμικού-EP.



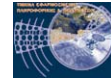
Σχ.2 Παρατηρήστε τις διαφορές των EPs από διάφορες (σε latency και τάση) συνιστώσες. Χρήση ICA για την εξαγωγή σημαντικής πληροφορίας.



Όπως φαίνεται και στο σχήμα τα EPs δεν μπορούν να εξεταστούν από τα αρχικά σήματα, αλλά πρέπει να εξαχθούν από κάποια μέθοδο (βλ. Σχ.2). Χρησιμοποιώντας το μέσο των σημάτων υπάρχει το πρόβλημα να χαθεί κάποια σημαντική πληροφορία. Αυτό οφείλεται στο γεγονός ότι το πλάτος της τάσης διαφέρει από δοκιμή σε δοκιμή, και εξαρτάται από την νοητική κατάσταση του ασθενούς την εκάστοτε χρονική στιγμή.

Προσπάθειες για την εξαγωγή των EPs έχουν γίνει με διάφορες τεχνικές επεξεργασίας σήματος. Μερικές από αυτές σχετίζονται με την εξομοίωση του σήματος, κάτι που επίσης δεν είναι ικανό να προσεγγίσει ρεαλιστικές τάσεις ή σχήματα των EPs και EEGs. Πλέον σήμερα χρησιμοποιούνται μέθοδοι σχετικά με τον υπολογισμό των άγνωστων πηγών, οι οποίες μεταδίδονται και αναμειγνύονται γραμμικά εντός ενός άγνωστου συστήματος το οποίο παράγει μείξεις σημάτων.

Αυτό το πρόβλημα μπορεί να λυθεί υπολογίζοντας τις πηγές με εφαρμογή BSS βασισμένη στην τεχνική ICA. Αυτή η μέθοδος μπορεί να εντοπίσει τα EPs από ανεξάρτητες πειραματικές δοκιμές και να διαχωρίσει τα EPs μέσα από κάθε συνιστώσα-IC. Οι πηγές των EPs και οι EEG συνιστώσες εκτιμώνται ως πολλαπλασιασμοί στα ηλεκτρόδια. Οι πολ/σμοι αυτοί προέρχονται από τις αυθεντικές πηγές, όπου και οι επιδράσεις όλων αυτών αθροίζονται. Κατά τη διάρκεια της εκπομπής τους, οι πηγές, αναμειγνύονται γραμμικά στο ηλεκτρικά άγνωστο σύστημα του εγκεφάλου, με αποτέλεσμα οι καταγραφές των ηλεκτροδίων αν είναι μείξεις διαφορετικών σημάτων.

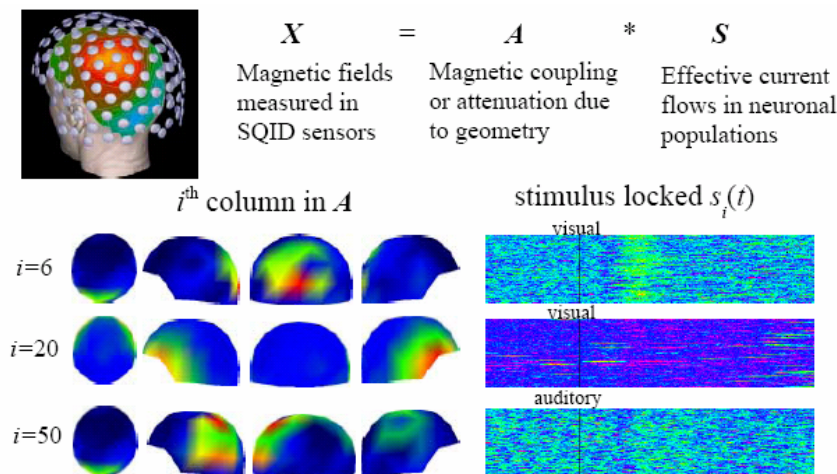


3 Το θεωρητικό υπόβαθρο της εργασίας

Το θεμελιώδες πρόβλημα στην έρευνα της αναγνώρισης προτύπων, καθώς επίσης και σε πολλές άλλες περιπτώσεις, είναι η κατάλληλη αναπαράσταση δεδομένων πολλών μεταβλητών, π.χ. τυχαία διανύσματα. Για λόγους ευκολίας υπολογισμού και θεμελίωσης, αυτή η αναπαράσταση συχνά αναζητείται ως ένας γραμμικός μετασχηματισμός των αυθεντικών δεδομένων (του ακατέργαστου διανύσματος). Γνωστές τεχνικές γραμμικών μετασχηματισμών είναι η Ανάλυση Κύριων Συνιστωσών (Principal Component Analysis – PCA), ICA, η Ανάλυση Συνιστωσών (Factor Analysis-FA) και η Επιδίωξη Προβολής (Projection Pursuit – PP). Οι τεχνικές αυτές βασίζονται στην BSS η οποία σαν σκοπό έχει να ανακτήσει ένα σύνολο από ανεξέλεγκτα σήματα ή πηγές με τη χρήση ενός μόνο ελεγχόμενου πίνακα.

4.1 Διαχωρισμός Πηγών στα Τυφλά (BSS)

Πρακτικά, οι παρατηρήσεις αφορούν την έξοδο ενός πίνακα από τους αισθητήρες, όπου κάθε ένας τους λαμβάνει ένα διαφορετικό συνδυασμό πηγών (βλ. Σχ.3).



Σχ.3 Η BSS καλείται να δώσει λύση στο πρόβλημα των μεικτών σημάτων που καταγράφονται στην επιφάνεια του κρανίου από τα ηλεκτρόδια. Τα σήματα των ηλεκτροδίων είναι γραμμικές αναπαραστάσεις από επιρροές διαφόρων πηγών.

Η πιο γενική διατύπωση για το πρόβλημα της BSS είναι ότι δεν υποθέτει τίποτα για την αρχική πληροφορία (τις πηγές) ή για τη διαδικασία μείξης.



Για την επίλυση αυτού του μοντέλου πρέπει να υπερβούμε την έλλειψη της αρχικής πληροφορίας θέτουμε τα εξής:

1. Οι πηγές είναι στατιστικά ανεξάρτητες. Αυτός είναι στατιστικά αυστηρό κριτήριο αλλά και μία λογική υπόθεση.
2. Το μοντέλο μείξης να είναι γραμμικό και στιγμιαίο.

Θεωρούμε $\mathbf{s} = [s_1(t), \dots, s_n(t)]^T$ με n αρχικές άγνωστες πηγές και $\mathbf{x} = [x_1(t), \dots, x_m(t)]^T$ με m γνωστές γραμμικές και στιγμιαίες μείξεις.

Το T υποδειλώνει αντιμετάθεση.

Το μοντέλο BSS εκφράζεται ως εξής:

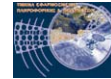
$$\mathbf{x}(t) = \sum_{j=1}^m \mathbf{a}_j s_j(t) = \mathbf{A}\mathbf{s}(t) \quad (1)$$

όπου $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_m]$ είναι ένας άγνωστος $n \times m$ πίνακας. Το πρόβλημα του BSS σχετίζεται με την προσεγγιση του πίνακα διαχωρισμού $\mathbf{B} \approx \mathbf{A}^{-1}$, η διαδικασία μείξης \mathbf{A} μπορεί να αντιστραφεί, με αποτέλεσμα πηγές \mathbf{s} να ανακτηθούν:

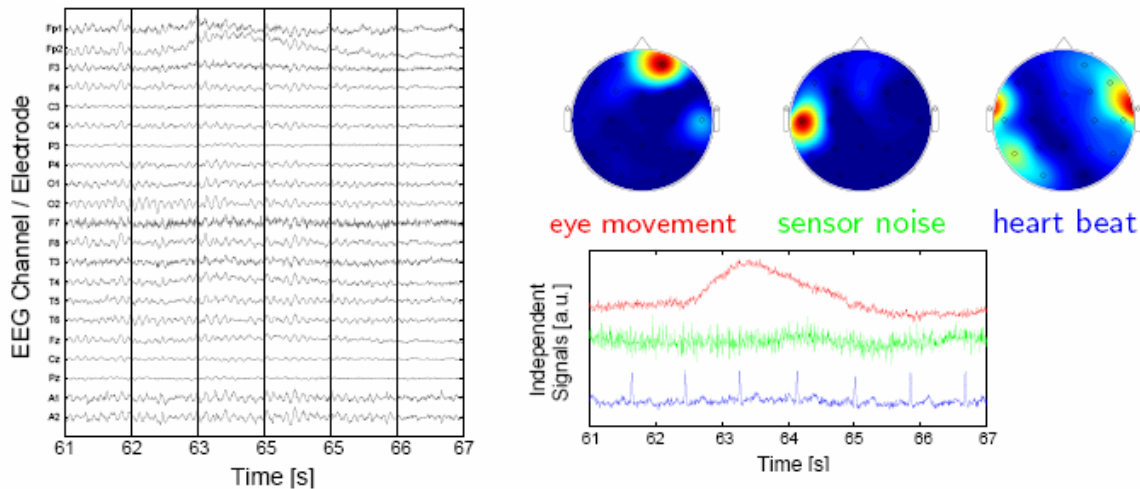
$$\mathbf{y} = \hat{\mathbf{s}} = \mathbf{B}\mathbf{x} = \mathbf{B}\mathbf{A}\mathbf{s} \approx \mathbf{s} \quad (2)$$

Το γενικό πρόβλημα της BSS απαιτεί ο \mathbf{A} να είναι πίνακας διαστάσεων $m \times n$ σε πλήρη διαβάθμιση, με $n \geq m$ (δηλ. να υπάρχουν τουλάχιστον τόσες μείξεις όσες και οι πηγές).

Κάτω από αυτές τις συνθήκες, οι πηγές προσδιορίζονται με μοναδικό τρόπο από τις μείξεις τους, και ως προς την αναδιαρθωσή και ως προς την αντιμετάθεση τους, δηλ. $\mathbf{B}\mathbf{A} = \mathbf{P}\mathbf{D}$, όπου \mathbf{D} είναι ένας διαγώνιος πίνακας και \mathbf{P} ο πίνακας αντιμετάθεσης.



Όπως είναι φυσικό στα πλαίσια της BSS, θα υποθέσουμε για τη συνέχεια ότι το πλήθος των πηγών είναι ίσο με τα ηλεκτρόδια, δηλ. $n = m$.



Σχ.4 Ο διαχωρισμός πηγών απομονώνει διαφορετικά είδη παρασιτικών σημάτων ή πηγών

Μία πιο γενική προσέγγιση θεωρεί πρόσθετο θόρυβο (βλ. Σχ.4), αλλά δεν θα αναφερθούμε σε αυτή.

Μια επιπλέον συνηθισμένη υπόθεση είναι ότι κάθε πηγή είναι μία Ανεξάρτητη και Πανομοιότυπη Κατανομημένη (Α.Π.Δ.) διαδικασία¹. Με αυτή την υπόθεση δεν αναγκαζόμαστε πλέον να συμφωνούμε με το χρόνο και θεωρούμε κάθε πηγή απλά σαν μία τυχαία μεταβλητή.

¹ Η Α.Π.Δ. είναι μία στοχαστική διαδικασία, αν τα στατιστικά της διαδικασίας δεν εξαρτώνται από τη ένδειξη του χρόνου. Η συνάρτηση αυτοσυσχέτισης μίας Α.Π.Δ.

$x(t)$ είναι $R_x(\tau) = (m_x - \sigma_x^2) \delta(\tau)$ όπου m_x είναι ο μέσος και σ_x^2 η διακύμανση της διαδικασίας. Το $\delta(\tau)$ είναι το «δ» της συνάρτησης Dirac. Μία Α.Π.Δ. με μηδενικό μέσο είναι μία λευκή διαδικασία (γενικά, το αντίστροφο δεν ισχύει).

4.1.1 Χωρική BSS

Οι αλγόριθμοι αυτής της κατηγορίας υποθέτουν ότι οι πηγές απορρίπτουν κάθε χρονική δομή από τα δεδομένα. Αν τα δεδομένα, όμως, έχουν ισχυρούς αυτοσυσχετισμούς τα αποτελέσματα αυτών των μεθόδων είναι αδύνατα [31], [37].



Για να υπολογίσουμε, λοιπόν, τον πίνακα διαχωρισμού οι αλγόριθμοι προσπαθούν να ελαχιστοποιήσουν την εξάρτηση μεταξύ των συνιστωσών ενός διανύσματος y . Κύρια διαφορά μεταξύ των αλγόριθμων αποτελεί η επιλογή του μετρου εξάρτησης (Κοινή Πληροφορία – MI, Οριακή Εντροπία, Αθροιστές, κτλ.) και η τεχνική βελτιστοποίησης (ανυσματική ή αλγεβρική μέθοδο).

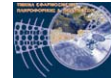
4.1.2 Χρονική BSS

Τα σήματα από τα εγκεφαλογραφήματα παρουσιάζουν σημαντική χρονική (καθώς και φασματική) δομή. Οι αλγόριθμοι BSS που βασίζονται στην ICA απορρίπτουν τελείως αυτή τη χρονική δομή και διαχωρίζουν τις πηγές εκμεταλευόμενοι την πολύ ισχυρή στατιστική υπόθεση περί ανεξαρτησίας μεταξύ των πηγών.

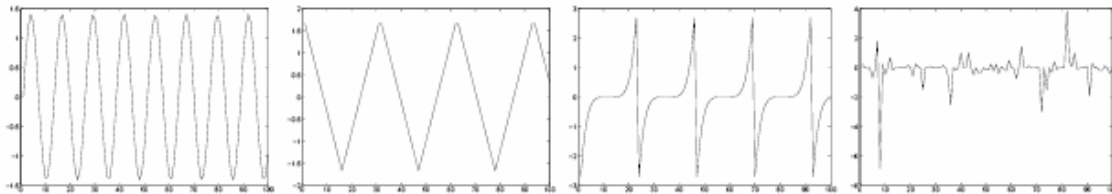
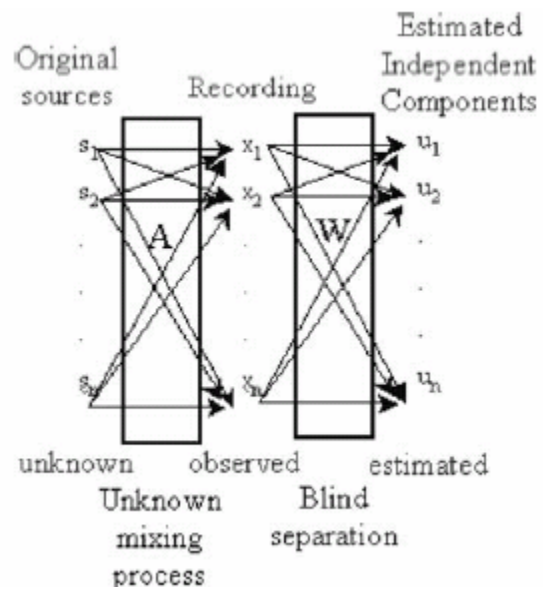
Αν όμως χρησιμοποιήσουμε τη χρονική και φασματική δομή των πηγών για να πετύχουμε το διαχωρισμό μπορούμε να είμαστε πιο ελαστικοί στην υπόθεση περί ανεξαρτησίας. Επιπλέον, η χωρικές προσεγγίσεις της BSS απαιτούν τον υπολογισμό πολύπλοκων στατιστικών μέτρων (οπως, MI ή Εντοπία), απεναντίας η χρονική BSS μπορεί να επιτευχθεί με χρήση μόνο απλών στατιστικών μέτρων δευτέρου βαθμού.

4.2 Ανάλυση Ανεξάρτητων Συνιστωσών (ICA)

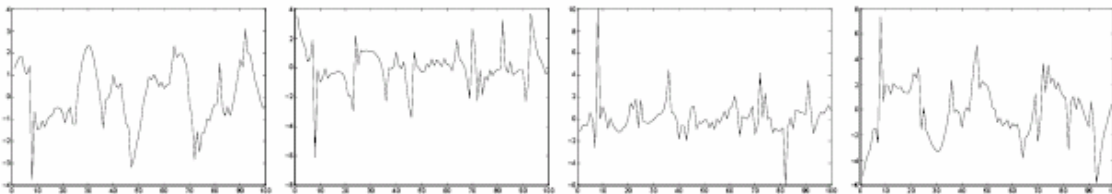
Ορίζοντας την ICA ενός τυχαίου $\mathbf{x} = [x_1, \dots, x_n]^T$ ως ένα αντιστρέψιμο μετασχηματισμό $\mathbf{y} = \mathbf{V}\mathbf{x}$, όπου \mathbf{V} είναι ένας πίνακας που εξασφαλίζει ότι η εξάρτηση μεταξύ των στοιχείων του μετασχηματισμένου διανύσματος $\mathbf{y} = [y_1, \dots, y_n]^T$ θα ελαχιστοποιηθεί.



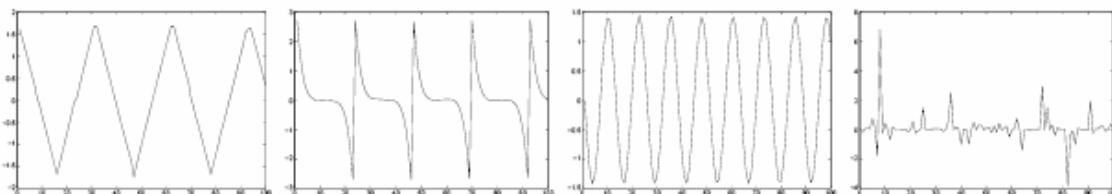
Σχ.5 Το σχήμα παρουσιάζει τη διαδικασία της ICA (δεξιά). Ακολουθεί ένα παράδειγμα με σχετικά σήματα της διαδικασίας. Οι αυθεντικές πηγές αναμιγνύονται σε ένα άγνωστο περιβάλλον και ανακτώνται με τη χρήση της ICA (κάτω).



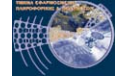
a) Αρχικές πηγές s .



b) Μείξεις $\mathbf{x} = \mathbf{A} \cdot \mathbf{s}$



c) ICs $\mathbf{u} = \mathbf{W} \cdot \mathbf{x}$



Η πίνακας \mathbf{V} της ICA προβολής είναι ο πίνακας διαχωρισμού $\mathbf{V} = \mathbf{B} = \hat{\mathbf{A}}^{-1}$ που αναζητούμε. Για να προσεγγίσουμε τον \mathbf{V} πίνακα προβολής πρέπει να καθορίσουμε ένα κριτήριο εξαρτησης(ανεξαρτησία) ή μία συνάρτηση αντίθεσης και ελαχιστοποίησης(μεγιστοποίηση) σε σχέση με τον \mathbf{V} .

Οι περισσότερες ICA συναρτήσεις αντίθεσης μπορεί να παραχθούν από την Αρχή Μέγιστης Πιθανότητας(ML). Αν υποθέσουμε ότι η Συναρτηση Πυκνότητας Πιθανότητας(PDF) κάθε πηγής s_i είναι $p_{s_i}(\cdot)$, τότε η PDF μίας πηγής τυχαίου διανύσματος $\mathbf{s} = [s_1, \dots, s_n]^T$ δίνεται από τη σχέση $P_s = \prod_{i=1}^n p_{s_i}(s_i)$, και η PDF των στοιχείων είναι για γνωστό πίνακα μείξης \mathbf{A} , ισούται με:

$$P(\mathbf{x}; \mathbf{A}, P_s) = |\mathbf{A}|^{-1} P_s(\mathbf{A}^{-1}\mathbf{x}) \quad (4)$$

όπου \mathbf{A}^{-1} είναι ο πίνακας διαχωρισμού τον οποίο θέλουμε να υπολογίσουμε. Οπότε, για ένα δοθέν σύνολο από T φορές πραγματοποίησης του \mathbf{x} , η κανονικοποιημένη λογαριθμική πιθανότητα δίνεται από τη σχέση:

$$L_T(\mathbf{A}, P_s) = \frac{1}{T} \sum_{t=1}^T \log P(\mathbf{x}(t); \mathbf{A}, P_s) \quad (5)$$

και η προσέγγιση της ML του πίνακα μείξης είναι $\mathbf{A}_{ML} = \arg \max_{\mathbf{A}} (L_T(\mathbf{A}, P_s))$, δηλ. η \mathbf{A}_{ML} έχει υπολογιστεί έτσι ώστε να μεγιστοποιηθεί η πιθανότητα εμφάνισης των δωθέντων δειγμάτων.

Για μεγάλο T , μπορεί να αποδειχτεί ότι [10]:

$$\lim_{T \rightarrow \infty} L_T(\mathbf{A}, P_s) = -H(P_x) - KL(P_y | P_s) \quad (6)$$



όπου το H δηλώνει τη διαφορική εντροπία, η KL είναι η απόκλιση Kullback-Leibler και $\mathbf{y} = \mathbf{A}^{-1}\mathbf{x}$. Η διαφορική εντροπία ενός τυχαίου διανύσματος $\mathbf{x} = [x_1, \dots, x_n]^T$ με πυκνότητα $f(\mathbf{x})$ δίνεται από:

$$H(f) = \int_{\mathbf{x}} f(\mathbf{x}) \log f(\mathbf{x}) d\mathbf{x} \quad (7)$$

Η απόκλιση KL μεταξύ δύο συναρτήσεων πυκνότητας πιθανότητας $f(\mathbf{u})$ και $g(\mathbf{u})$ δίνεται από:

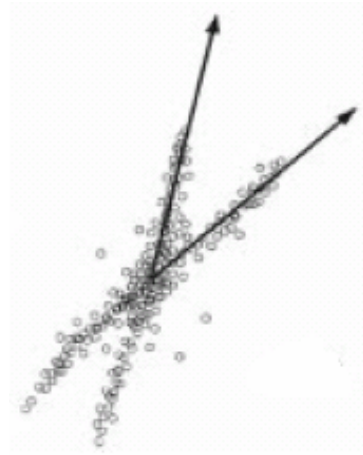
$$KL(f | g) = \int_{\mathbf{u}} f(\mathbf{u}) \log \frac{f(\mathbf{u})}{g(\mathbf{u})} d\mathbf{u} \quad (8)$$

Ο όρος $H(P_{\mathbf{x}})$ στην ισότητα 6 προσαρμόστηκε για κάποιες γνωστές μείξεις \mathbf{x} και δεν εξαρτάται από τις παραμέτρους του μοντέλου, δηλ. δεν εξαρτάται από τον άγνωστο πίνακα μείξης \mathbf{A} . Οπότε, η συνάρτηση αντίθεσης, την οποία πρέπει να ελαχιστοποιήσουμε για να προσεγγίσουμε τη ML του πίνακα μείξης, είναι:

$$\phi^{ML}(\mathbf{y}) = KL(P_{\mathbf{y}} | P_{\mathbf{s}}) \quad (9)$$

Επειδή η απόκλιση KL μπορεί να γίνει κατανοητή σαν το μέτρο της απόστασης μεταξύ των κατανομών η ερμηνεία για την ϕ^{ML} είναι η εξής: ο ιδανικός πίνακας διαχωρισμού $\mathbf{B} = \hat{\mathbf{A}}^{-1}$ βρίσκεται όταν η κατανομή του διανύσματος \mathbf{y} είναι όσο το δυνατό πιο κοντά στην κατανομή των πηγών \mathbf{s} .

Σχ.6 Η ICA για μία μη-ορθογωνική μείξη δύο κατανομών που είναι ανεξάρτητες και αξερατικά αραιωμένες (κορυφόμενες σε μεγάλη διακύμανση). Τα διανύσματα-διαχωρισμού παριστάνουν τη λύση της ICA. Είναι φανερό ότι η λύση της ICA καλύπτει τη μετασχηματισμένη μονάδα διανυσμάτων. Η χρησιμότητα του μη-ορθογωνικού μετασχηματισμού ευαίσθητου σε υψηλού βαθμού στατιστική, η οποία δείχνει τη θεμελιώδη συνάρτηση της ICA για μία δισδιάστατη (2-D) και non-Gaussian κατανομή. Όμοιες κατανομές είναι πρακτικά πιο συνηθισμένες σε πραγματικά δεδομένα παρά σε αυτές τις υποτειθέμενες μοντελοποιημένες "Gaussian μείξεις". Το γεγονός αυτό προέρχεται από την κοινή μη-ορθογωνική "μείξη" των ιδιαίτερα αραιών ανεξάρτητων συνιστωσών. Αραιό, εννοούμε μία κατανομή που περιέχει κορυφώσεις μεγάλης διακύμανσης (όχι μηδενικές, μεγάλες κορυφές) το αντίθετο από τις Gaussian κατανομές. Το αραιό είναι super-Gaussian, και συνήθως εντοπίζεται από την θετική κύρτωση.





Οι περισσότεροι ICA αλγόριθμοι κατατάσσονται σε 5 κύριες ομάδες σύμφωνα με την υπόθεση που θέτουμε για το P_s :

1. Μέθοδοι βασισμένες στην ML. Οι τεχνικές ICA αυτής της κατηγορίας υποθέτουν ένα σταθερό P_s και ελαχιστοποιούν το ϕ^{ML} μόνο μέσω του P_y . Αυτό σημαίνει ότι αυτές οι μέθοδοι επιλέγουν ένα πίνακα διαχωρισμού \mathbf{B} τέτοιον ώστε η κατανομή του $\mathbf{y} = \mathbf{B}\mathbf{x}$ να είναι όσο το δυνατό πιο κοντά (σύμφωνα με τις αρχές της απόκλισης KL) με την υποτειθέμενη κατανομή των πηγών.

Ο πιο γνωστός αλγόριθμος της κατηγορίας είναι ο InfoMax [3].

2. Μέθοδοι βασισμένες στην Κοινή Πληροφορία (MI). Οι τεχνικές ICA αυτής της κατηγορίας δεν υποθέτουν κάτι σχετικά με την κατανομή των πηγών και προσπαθούν να ελαχιστοποιήσουν το ϕ^{ML} όχι μόνο μέσω του \mathbf{B} αλλά λαμβάνοντας επίσης υπόψη το P_s . Αυτή η ελαχιστοποίηση αποφέρει τη αντιστροφή συνάρτηση της ML [10]:

$$\phi^{MI}(\mathbf{y}) = KL\left(P_y \mid \prod_{i=1}^n p_{y_i}(y_i)\right) = MI(P_y) \quad (10)$$

όπου το $p_{y_i}(\cdot)$ είναι η οριακή κατανομή της ανεξάρτητης i -στης συνιστώσας και η $MI(P_y)$ υποδηλώνει την κοινή πληροφορία μεταξύ των συνιστωσών του τυχαίου διανύσματος \mathbf{y} .

Η $MI(P_y)$ είναι ένα μέτρο εξάρτησης μεταξύ των στοιχείων του τυχαίου διανύσματος \mathbf{y} . Πάντα είναι αρνητικό, και μηδέν μόνο αν τα στοιχεία του \mathbf{y} είναι ανεξάρτητα.

Άρα, αν έχουμε στη διαθεσή μας μία αξιόπιστη προσέγγιση της MI , μπορούμε να εντοπίσουμε τον πίνακα διαχωρισμού \mathbf{B} ελαχιστοποιώντας τη

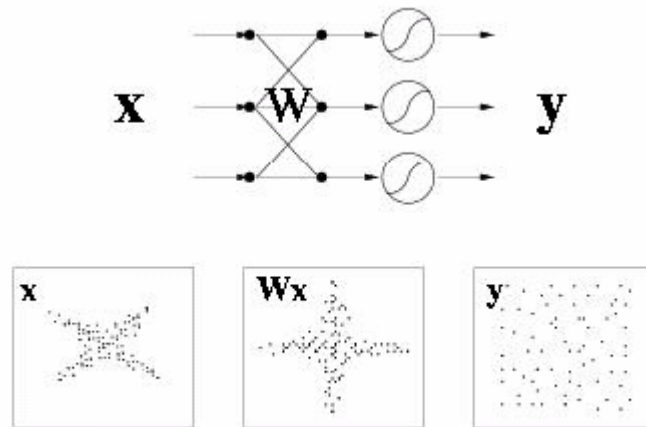


MI μεταξύ των συνιστωσών του τυχαίου διανύσματος $\mathbf{y} = \mathbf{B}\mathbf{x}$. Αυτό επιτυγχάνουν οι αλγόριθμοι $MILCA$ [38] και $NPICA$ [7].

3. Μέθοδοι βασισμένες στην Οριακή Εντροπία. Σε αντίθεση με την κατηγορία της MI , δεν υποθέτουν τίποτα σχετικά με το P_s .

Οστόσο, υποθέτουν ότι, όταν οι συνιστώσες του \mathbf{y} είναι ανεξάρτητες, είναι και ασυσχέτιστες, δηλ. υποθέτουν δευτέρου βαθμού ανεξαρτησία [8].

Αν έχουμε μόνο ένα δείγμα πεπερασμένου μεγέθους, τότε έχουμε μικρότερα βήματα στην επίδοση της διαδικασίας διαχωρισμού, λόγω λαθών στην προσέγγιση του $\hat{\Sigma}_x$, όπου $\hat{\Sigma}_x$ είναι η προσέγγιση του πίνακα της συνδιακύμανσης των δειγμάτων.



Σχ.7 Η ανεξαρτησία και η μέγιστη εντροπία. Με την μεγιστοποίηση της εντροπίας, έπειτα από ένα μη-γραμμικό μετασχηματισμό, δίνει ανεξάρτητες συνιστώσες.

Υπό την υπόθεση του μη συσχετισμού, ακολουθούμε 2 βήματα για την εύρεση του πίνακα διαχωρισμού \mathbf{B} :

- Ο υπολογισμός του πίνακα μη συσχετισμού (ή λευκανσης) $\mathbf{W} = \hat{\Sigma}_x^{-1/2}$ για τα εξεταζόμενα δείγματα των δεδομένων μας.
- Η εύρεση του ορθογώνιου πίνακα \mathbf{R} με την ελαχιστοποίηση του $\phi^{MI}(\mathbf{R}\mathbf{W}\mathbf{x})$.



Η ελαχιστοποίηση του ϕ^{MI} λαμβάνοντας υπόψη τον ορθογώνιο πίνακα² είναι το ίδιο με το να ελαχιστοποιηθεί η εξής αντίστροφη συναρτηση:

$$\phi^{ME}(\mathbf{y}) = \sum_{i=1}^n H(p_{y_i}) \quad (11)$$

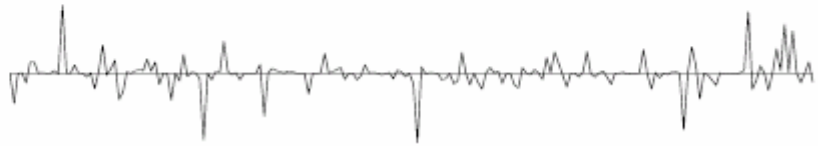
όπου το $H(p_{y_i})$ είναι η Οριακή Εντροπία των υπολογισθέντων αναξάρτητων συνιστωσών (ICs) y_i .

Ελαχιστοποιώντας την εντροπία των εκτιμώμενων πηγών είναι το ίδιο με το να κάνουμε τις πηγές όσο πιο μη-Γκαουσιανες (non-gaussian) γίνεται. Γνωστοί αλγόριθμοι της κατηγορίας είναι ο *FastICA* [23] και *RADICAL* [26].

gaussian:



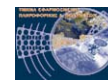
sparse:



Σχ.8 Η αραιότητα είναι μία μορφή non-Gaussianity (υψηλού βαθμού μορφή) και συχνά συναντάται σε φυσικά σήματα. Η διακύμανση(ενέργεια) εμφανίζεται σπάνια.

4. Μέθοδοι βασισμένες σε Υψηλού Βαθμού Στατιστική (HOS). Αυτές οι μέθοδοι βασικά υποθέτουν ότι η κατανομή των πηγών είναι κοντά στο επίπεδο του φυσιολογικού. Υπό αυτή την προϋπόθεση οι Συναρτήσεις Πυκνότητας Κατανομής των πηγών μπορούν να προσεγγιστούν σύμφωνα με τα αθροίσματα (κυριώς της κύρτωσης και της διακύμανσης), έτσι μπορεί να υπολογιστούν τα ϕ^{ME} και ϕ^{MI} .

Μέσω της ελαχιστοποίησης των εκτιμώμενων αντιστροφών συναρτήσεων υπολογίζεται ο πίνακας διαχωρισμού. Αυτό ακριβώς επιτυγχάνει ο αλγόριθμος *JADE* [10]. Μία πιά γενική προσέγγιση βασισμένη σε αυτή την αρχή είναι ο *NICA* [35].



5. Μέθοδοι βασισμένες στο Συσχετισμό. Αυτές οι μέθοδοι χρησιμοποιούν μη-γραμμικούς συσχετισμούς μεταξύ των εκτιμώμενων ανεξάρτητων συνιστωσών(ICs) για να υπολογιστεί η ΜΙ τους. Ένας πολύ ικανός αλγόριθμος αυτής της κατηγορίας είναι ο *KICA* [2].
6. Επίσης, υπάρχει μία ακόμη κατηγορία βασισμένη στη θεωρία Bayes, αν και δεν θα αναφερθούν παραπάνω πληροφορίες, υπογραμμίζεται ότι υπάρχει ιδιαίτερο ενδιαφέρον σχετικά με αυτή.

4.3 Εφαρμόζοντας την ICA σε δεδομένα EEG

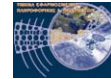
Με σκοπό η ICA να είναι ικανή να διαχωρίσει τις πηγές υποθέτουμε τα παρακάτω:

1. Οι πηγές είναι στατιστικά ανεξάρτητες.
2. Η διαδικασία μείξης είναι γραμμική και στιγμιαία.
3. Κάθε πηγή να μπορεί να αναπαρασταθεί.
4. Οι πηγές να μην έχουν Gaussian Συναρτήσεις Πυκνότητας Πιθανότητας. Οι πηγές δεν μπορούν να είναι Gaussian (εκτός της μονάδας) επειδή η ανεξαρτησία είναι μη συσχετιζόμενη για Gaussian κατανομές.

Αυτό σημαίνει ότι αν το s είναι τυχαίο διάνυσμα με ανεξάρτητες Gaussian συνιστώσες και $\mathbf{x} = \mathbf{A}s$ είναι οι μείξεις, τότε ο πίνακας διαχωρισμού που δίνει ανεξάρτητες συνιστώσες είναι ο:

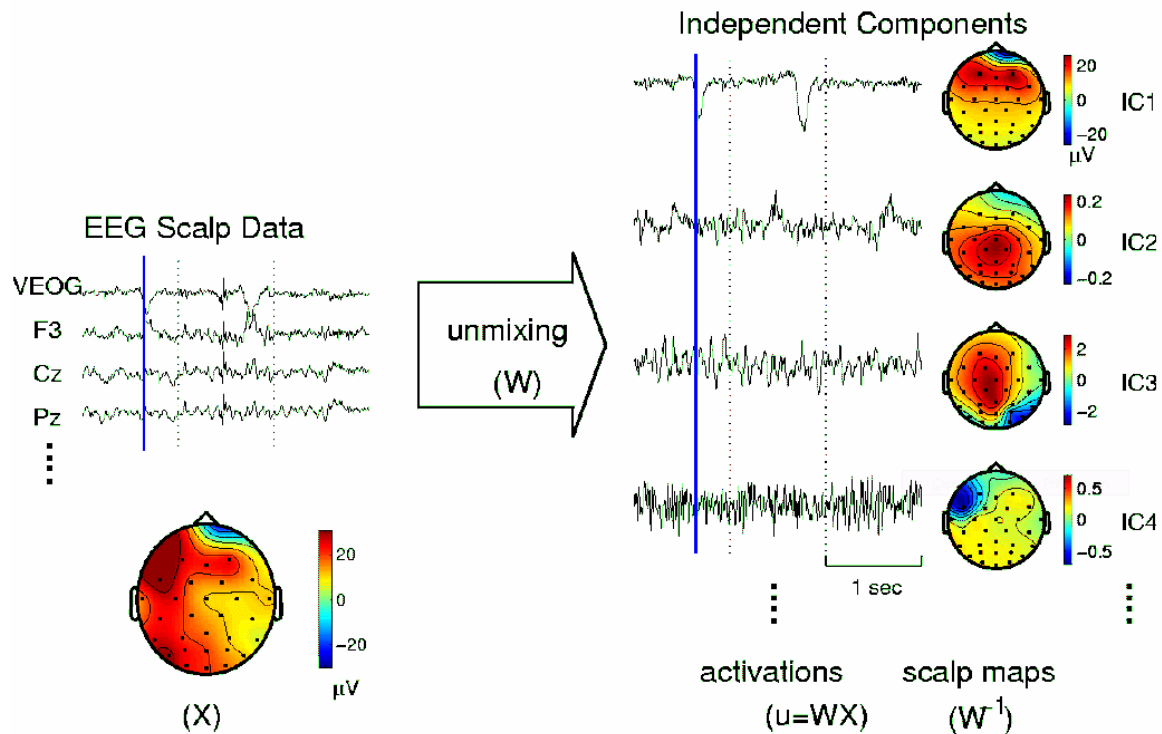
$$\mathbf{B} = \mathbf{R}\hat{\Sigma}_{\mathbf{x}}^{-1/2} \quad (12)$$

όπου $\hat{\Sigma}_{\mathbf{x}}$ είναι ο εκτιμώμενος πίνακας διασποράς των στοιχείων και \mathbf{R} είναι κάθε ορθογώνιος πίνακας (ή αλλιώς πίνακα περιστροφής). Άρα, για την περίπτωση των περισσότερων του ενός Gaussian πηγών, ο



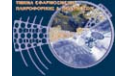
πίνακας διαχωρισμού \mathbf{B} μπορεί να καθοριστεί μόνο από ένα άγνωστο πίνακα περιστροφής \mathbf{R} .

ICA Decomposition into Independent Components



Σχ.9 Το σχεδιάγραμμα επάνω παρουσιάζει διαφωτιστική περιγραφή. Στην ανάλυση EEG, οι γράμμες του πίνακα εισαγωγής, \mathbf{X} , είναι τα σήματα EEG που έχουν καταγραφεί από τα ηλεκτρόδια και οι στήλες είναι οι μετρήσεις για κάθε χρονική στιγμή (scalp maps – κρανιακοί χάρτες). Η ICA υπολογίζει ένα πίνακας διαχωρισμού, \mathbf{W} , ο οποίος αποσυνθέτει ή διαχωρίζει γραμμικά τα πολυκαναλικά δεδομένα σε ένα σύνολο χρονικά ανεξάρτητων και χωρικά σταθερών συνιστωσών. Οι γραμμές του πίνακα εξόδου, $\mathbf{u} = \mathbf{W}\mathbf{X}$, είναι οι χρονοσειρές των ενεργοποιητών των ανεξάρτητων συνιστωσών. Οι στήλες του αντιστροφου πίνακα, $\text{inv}(\mathbf{W})$, δίνει τη σχετική ισχύ των προβολών των αντιστοιχών συνιστωσών για κάθε κρανιακό αισθητήρα. Αυτά τα κρανιακά βάρη σκιαγραφούν την κρανιακή τοπογραφία για κάθε συνιστώσα, και υποδεικνύει την φυσιολογική προέλευση των συνιστωσών.

Αν υποθέσουμε ότι η πολυπλοκότητα των EEG δυναμικών μπορεί να αναπαρασταθεί, τουλάχιστο ένα μέρος αυτών, ως μία συλλογή από ένα ικανοποιητικό αριθμό ανεξάρτητων στατιστικά εγκεφαλικών διαδικασιών, τότε η ανάλυση των EEG πηγών ικανοποιεί την υπόθεση (1).



Η υπόθεση (2) ικανοποιείται εφόσον η ένταση της αγωγιμότητας στο πλέγμα του κρανίου, για τις συχνότητες των EEG, είναι ικανοποιητικά στιγμιαία και σχεδόν γραμμική.

Η υπόθεση (3) δεν ικανοποιείται σαφώς διότι, οι εγκεφαλικές διαδικασίες είναι χρονοσειρές και δεν μπορούν να αναπαραστηθούν σαν διαδικασίες. Αυτό λοιπόν το γεγονός σχεδόν αγνοείται στις μελέτες των EEG με ICA. Παράλληλα, πρόσφατα αποδείχτηκε ότι, αν εμφανίζονται ισχυρές χρονικές εξαρτήσεις στα δεδομένα, τότε η προσέγγιση των πηγών από την ICA είναι ανέφικτη (βλέπε [31], [37]).

Αν και δεν γνωρίζουμε αν η υπόθεση (4) ικανοποιείται εξαιτίας του πεπερασμένου μεγέθους των δεδομένων η απόλυτη Gaussian κατανομή είναι απίθανη.

4.4 InfoMax

Ο αλγόριθμός υλοποιεί την τεχνική της ICA και ανήκει στην κατηγορία των χωρικών αλγορίθμων της BSS.

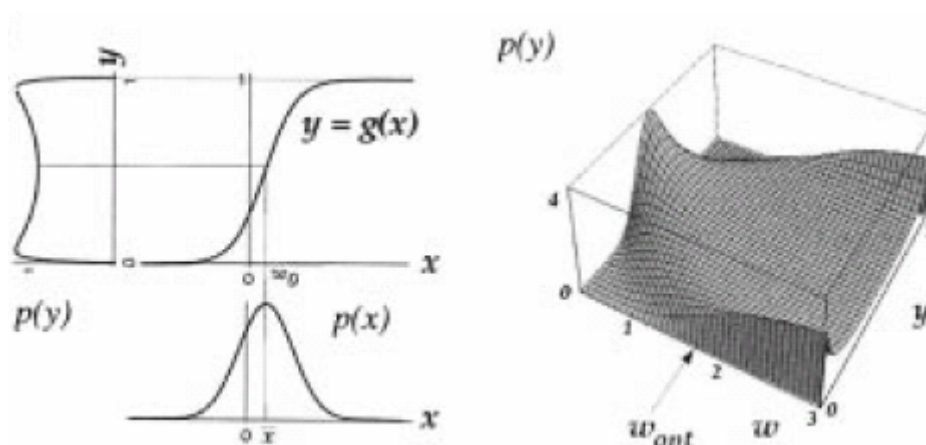
Κατα τη μεγιστοποίηση της λογαριθμικής συνάρτησης πιθανότητας λαμβάνοντας υπόψη τη σχέση $\mathbf{W} = \mathbf{A}^{-1}$ και χρησιμοποιώντας στοχαστική ανυσματική κλίση καταλήγουμε στον παρακάτω αλγόριθμο μάθησης για το \mathbf{W} [3]:

$$\Delta \mathbf{W} \propto \left[(\mathbf{W}^T)^{-1} - \varphi(\mathbf{y}) \mathbf{x}^T \right] \quad (1)$$

όπου το $\mathbf{y} = \mathbf{W}\mathbf{x}$ είναι οι ανεξάρτητες συνιστώσες, το $\varphi(\mathbf{y}) = -\frac{\partial p(\mathbf{y})/\partial \mathbf{y}}{p(\mathbf{y})}$, και το $p(\mathbf{y}) = \prod_{i=1}^n p_{y_i}(y_i)$ είναι η υποτειθέμενη κατανομή των ανεξάρτητων συνιστωσών (δηλ. η κατανομή των πηγών). Ένας πιο αποτελεσματικός τρόπος να μεγιστοποιηθεί η λογαριθμική πιθανότητα είναι με το ακόλουθο «φυσικό» άνυσμα [1]:

$$\Delta \mathbf{W} \propto \left[\mathbf{I} - \varphi(\mathbf{y}) \mathbf{y}^T \right] \mathbf{W} \quad (2)$$

όπου I είναι ο ταυτόσημος πίνακας.



Σχ.10 Ένα διάγραμμα πληροφορίας σιγμοειδών νευρώνων. (αριστερά) Η είσοδος x με συνάρτηση πυκνότητας πιθανότητας $p(x)$, όπου το n είναι Gaussian, διέρχεται από μία μη γραμμική συνάρτηση $g(x)$. Η πληροφορία για την προκύπτουσα πυκνότητα, $p(x)$ εξαρτάται από τη σύγκλιση του μέσου και της διακύμανσης του x στο όριο, w , του $g(x)$ (Nicol Schraudolph, personal communication). (δεξιά) Το $p(y)$ είναι σχεδιασμένο για διάφορες τιμές βαρών w , όπου μία απλή μονάδα Infomax επιχειρεί να συκλίνει μία Gaussian κατανομή σε μία λογιστική σιγμοειδή μονάδα. Το ιδανικό βάρος, w μεταδίδει την πλειοψηφία της πληροφορίας (από το [14]). Το ταίριασμα δεν μπορεί να είναι τέλει, αλλά όμως καταφέρνει να προσεγγίσει την συνάρτηση πυκνότητας πιθανότητας της μέγιστης εντροπίας για τη μονάδα κατανομής μεγιστοποιώντας την αναμενόμενη λογαριθμική κλίση.

Η τοπική σύγκλιση εξασφαλίζεται μόνο αν το $p(y)$ είναι η πραγματική λογαριθμική πυκνότητα των πηγών [34].

Μία επεκταμένη μέθοδος του InfoMax είναι ικανή να διαχωρίσει εξίσου sub-Gaussian και super-Gaussian πηγές όπως αυτό εξηγείται στο [27].

Ο επεκταμένος InfoMax απαιτεί τον ορισμό των συνδεδειγμένων κριτηρίων μεταξύ των Συναρτήσεων Πυκνότητας Πιθανότητας και της Παραμετρικότητας.

Τα μειονεκτήματα του αλγόριθμου:

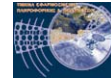
- Χαμηλή προσαρμοστικότητα στις κατανομές των πηγών
- Σχετικά χαμηλή ταχύτητα σύγκλισης για αλγεβρικές τεχνικές όπως JADE ή σταθερού σημείου όπως ο FastICA.



- Για τις ανυσματικές μεθόδους η τεχνική βελτιστοποίησης μπορεί να συγκλίνει στο τοπικό ελάχιστο της αντιστροφής συνάρτησης με αποτέλεσμα να μην επιτυγχάνει καλή εκτίμηση του πίνακα διαχωρισμού.

Τα πλεονεκτήματα:

- Η απλότητα του αλγόριθμου
- Οι δυνατότητα για υλοποίηση του από νευρωνικό δίκτυο, κάτι που απαιτεί μικρότερη υπολογιστική ισχύ.



4.5 Περιβάλλον υλοποίησης και πακέτα λογισμικού

Η συγκεκριμένη εργασία έχει υλοποιηθεί στο προγραμματιστικό πακέτο του MatLab σε Linux (για σταθερότητα και ταχύτητα εξομοίωσης). Το MatLab είναι ένα ισχυρό και ευρή πρόγραμμα το οποίο χρησιμοποιείται για τεχνικούς υπολογισμούς και οπτικοποίηση των αποτελεσμάτων. Στην ουσία το MatLab είναι μια συλλογή βιβλιοθηκών της σκριπτογλωσσας του πακετου.

Για την μελέτη των δεδομένων έχει χρησιμοποιηθεί και δοκιμαστεί κώδικας με αλγόριθμους και μεθοδολογίες από διαφορα πρωτοπόρα εργαστήρια που δραστηριοποιούνται στον τομέα. Κυρίως τα πακέτα που έχουν χρησιμοποιηθεί περισσότερο είναι το EEGLAB (Salk Institute, San Diego), FastICA (HUT, Finland) και ICAEPTtoolbox (Biopattern Research Group). Ο κώδικας αυτών των πακέτων έχει τροποποιηθεί ή έχουν υλοποιηθεί επιπλέον συναρτήσεις για την τελική ανάλυση των δεδομένων.

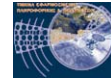
```

34 [features] = extractClusteringFeatures (allSubjectsComponents, t1, t2, srate, i);
35 h = plotFeatures (features, srate);
36
37 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% Primary Clustering of the ICA
38
39 [allSubjectsComponents, clusters] = getClusteredComponents (features, allSubjectsComponents, 'kmeans', 10);
40
41 [h1, h2, h3] = dispClusters_meanActivations (clusters, allSubjectsComponents, t1, t2, srate);
42 [h1, h2, h3] = dispClusters_meanBackProjections (clusters, allSubjectsComponents, t1, t2, srate);
43 h = dispClusters_activations (clusters, allSubjectsComponents, t1, t2, srate);
44 h = dispClusters_backProjections (clusters, allSubjectsComponents, t1, t2, srate);
45 h = dispClusters_colMixing (clusters, allSubjectsComponents);
46
47 for clusterNumber = 1:clusters.nClusters
48     displayCluster_image_activations (allSubjectsComponents, clusterNumber, t1, t2, srate);
49     displayCluster_image_backprojections (allSubjectsComponents, clusterNumber, t1, t2, srate);
50 end
51
52 for clusterNumber = 1:clusters.nClusters
53     CompIds = [];
54     [CompIds] = getOneClusterIndices (clusterNumber, allSubjectsComponents);
55     isAct = 1; % plotting activations
56     isTopo = 0; % not plotting topography
57     title = sprintf('cluster %d', clusterNumber);
58     display_ESPIMAGE (allSubjectsComponents(CompIds), t1, t2, srate, electrodesLoc, isAct, isTopo, title);
59 end
60
61 for clusterNumber = 1:clusters.nClusters
62     CompIds = [];
63     [CompIds] = getOneClusterIndices (clusterNumber, allSubjectsComponents);
64     isAct = 0; % plotting backprojections
65     isTopo = 1; % plotting topography
66     title = sprintf('cluster %d', clusterNumber);
67     display_ESPIMAGE (allSubjectsComponents(CompIds), t1, t2, srate, electrodesLoc, isAct, isTopo, title);
68 end
69
70 for clusterNumber = 1:clusters.nClusters
71     CompIds = [];
72     [CompIds] = getOneClusterIndices (clusterNumber, allSubjectsComponents);
73     title = sprintf('cluster %d', clusterNumber);
74     displayCluster_InstabilityOfA (allSubjectsComponents(CompIds), clusterNumber, t1, t2, srate, title);
75 end
76
77
78 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% Secondary Clustering of the ICA
79

```



Σχ.11 Το περιβάλλον του MatLab που τρέχει σε κέλυφος Linux(πάνω). Περιβάλλον για ανάπτυξη κώδικα(δίπλα).

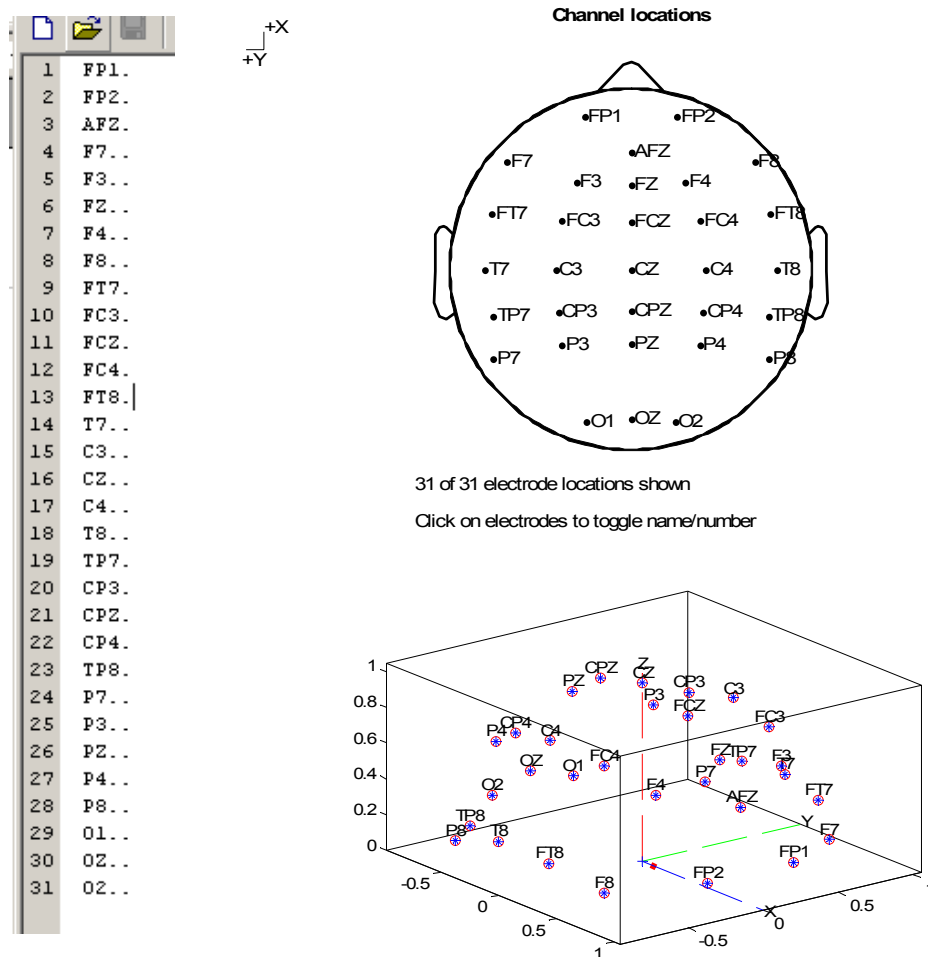


4 Πειραματική διαδικασία

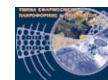
Η πειραματική διαδικασία που είναι μία δοκιμή μιας προτεινόμενης μεθοδολογίας, από τους συνεργάτες του ερευνητικού προγράμματος BIOPATTERN και για το οποίο εργαζομαι, πάνω σε τεχνητά δεδομένα.

4.1 Περιγραφή Δεδομένων

4.1.1 Location file (Αρχείο Ηλεκτροδίων)



Σχ.12 Στο σχήμα παρουσιάζεται το αρχείο των ηλεκτροδίων σε 2-D, 3-D και κατά σειρά κατάταξης των ηλεκτροδίων που αντιστοιχούν στις καταγεγραμμένες χρονοσειρές.

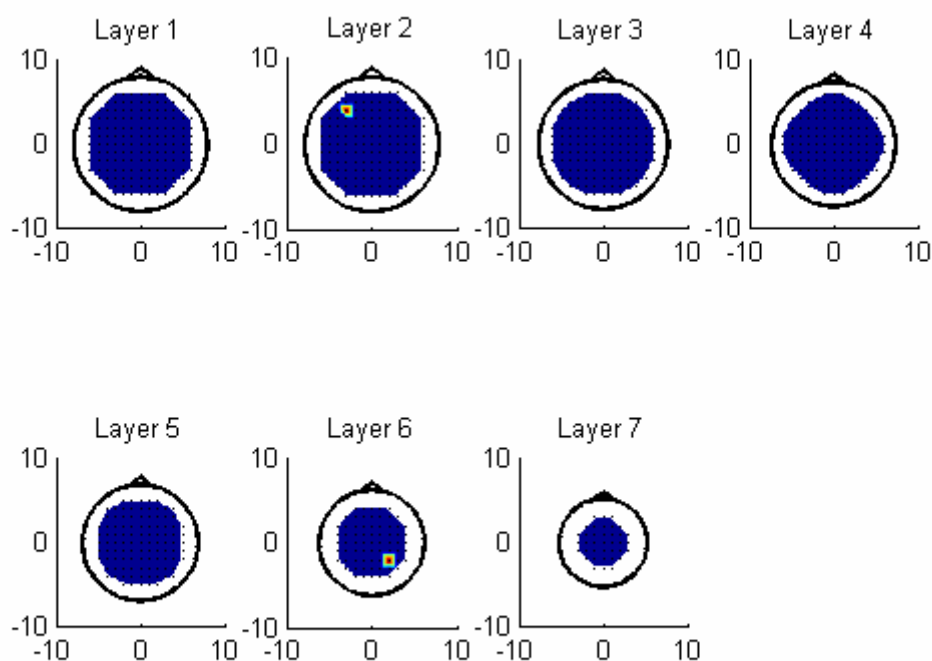


4.1.2 Τεχνητά σήματα EEG με γνωστές πηγές

Δύο διπολικές πηγές έχουν καθοριστεί στο 3^ο επίπεδο του σφαιρικού εγκεφαλικού μοντέλου όπως φαίνεται παρακάτω:

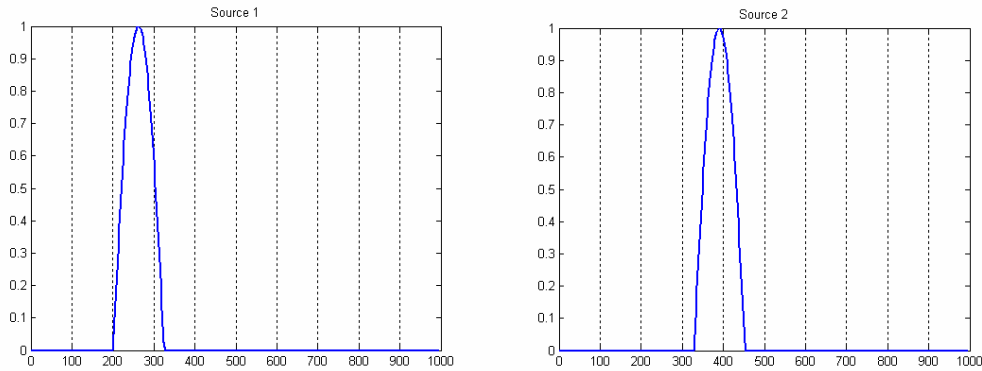
1^η πηγή: Τοποθετημένη στις (-3, 4, 1) σφαιρικές συντεταγμένες σε μία θετική κατεύθυνση του y

2^η πηγή: Τοποθετημένη στις (2, -2, 5) σφαιρικές συντεταγμένες σε μία θετική κατεύθυνση του y



Σχ.13 Παρουσίαση των πηγών με βάση των οποίων έχουν δημιουργηθεί τα σήματα του πειράματος.

Οι χρονοσειρές που έχουν δημιουργηθεί από τις δύο πηγές αποτελούνται αντίστοιχα από μισά ημίτονα των 4Hz με ρυθμό δειγματοληψίας 250Hz. Για λόγους απλότητας, οι πηγές δεν επικαλύπτονται και καλύπτουν το χρονικό διάστημα των [200-456] ms.



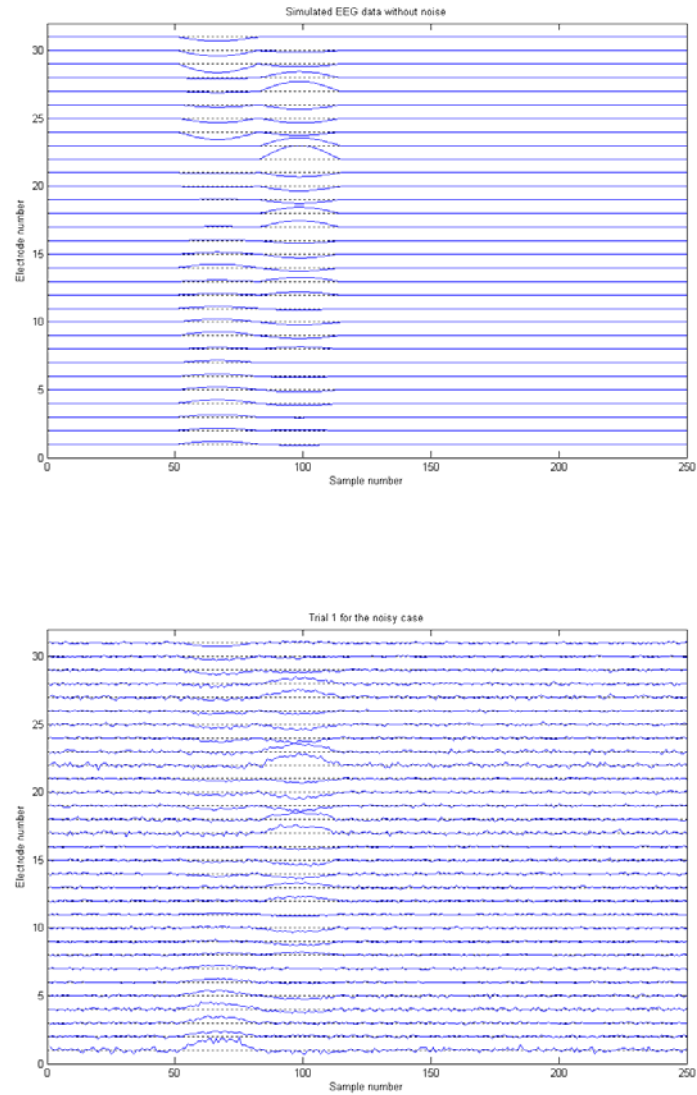
Σχ.14 Τα ημίτονα των 4Hz για τις δύο πηγές που καλύπτουν το χρονικό διάστημα των [200-456]ms.

Λύνοντας το πρόβλημα της προώθησης $\mathbf{F} = \mathbf{KJ}$, όπου το \mathbf{J} το τρέχον διπολικό διάνυσμα, το \mathbf{K} είναι το ο πίνακας απολαβής και το \mathbf{F} είναι ο πίνακας των κρανιακών δυναμικών. Κάθε σύνολο απο κρανιακές καταγραφές αποτελείται από 31 ηλεκτρόδια των 250 στιγμιαίων δειγμάτων(frames).

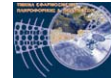
Θεωρούμε δύο περιβάλλοντα:

1. χωρίς θόρυβο
2. με θόρυβο

Και τα δύο περιβάλλοντα, εξομοιώνονται για 40 πειραματικές δοκιμές(trials). Στην περίπτωση θορύβου, η μόνη μεταβλητή σε όλα τα trials είναι ο λευκός Gaussian θόρυβος με SNR 5dB, όπου προστείνεται κατα μήκος των κρανιακών καταγραφών.

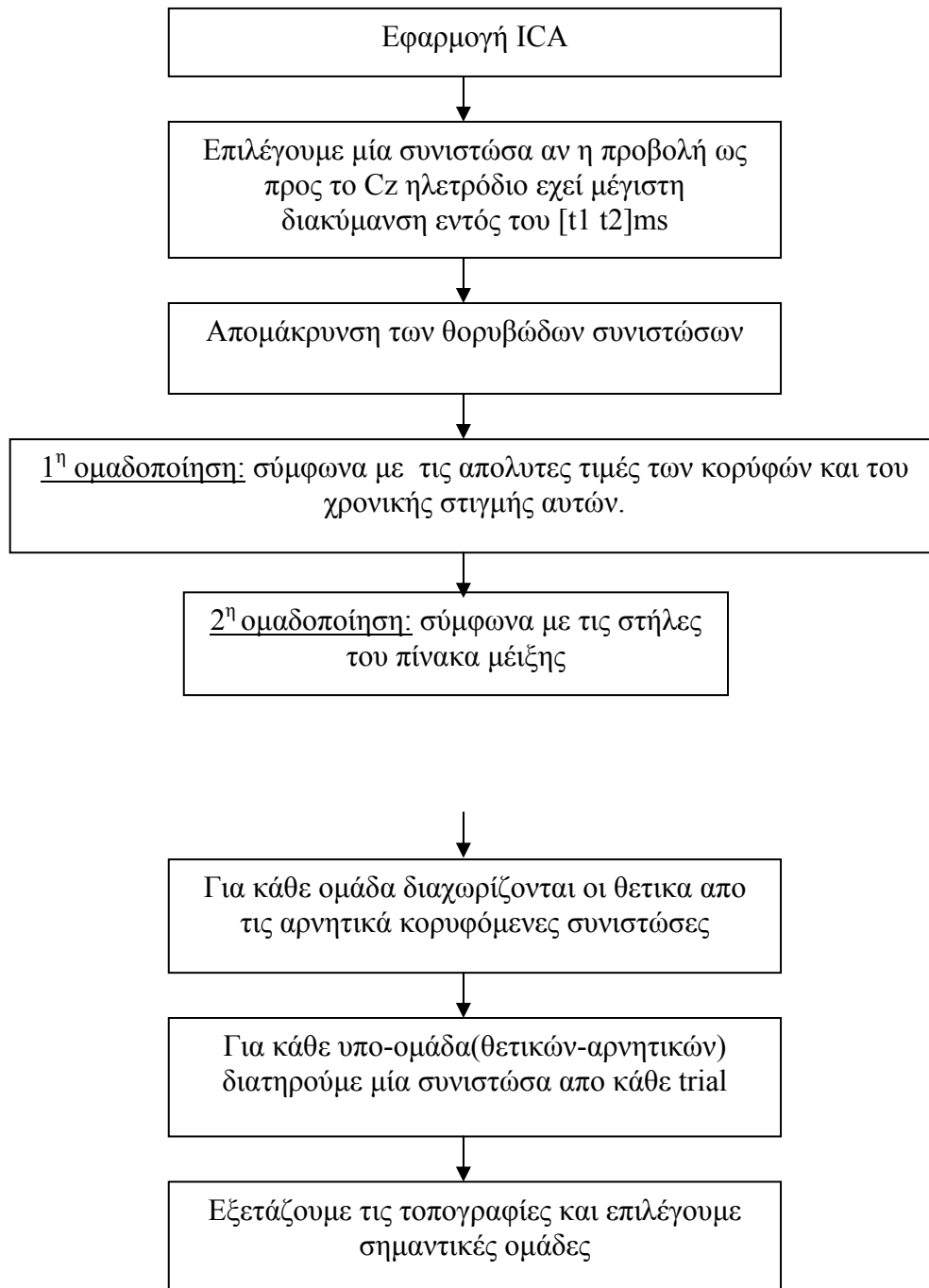


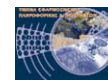
Σχ.15 Τα παρακάτω γραφήματα παρουσιάζουν τα εξομοιωμένα δυναμικά του 1^{ου} trial.



4.2 Μεθοδολογία ανάλυσης δεδομένων

Το παρακάτω σχεδιάγραμμα περιγράφει τη διαδικασία που εφαρμόστηκε για την εξαγωγή των αποτελεσμάτων.





4.2.1 Επιλογή Συνιστώσων (ICs)

4.2.1.1 Μέγιστη Διακύμανση

Η ICA εφαρμόζεται σε κάθε trial, στη συνέχεια επιλέγονται οι συνιστώσες που έχουν μέγιστες προβολές στο Cz εντός του χρονικού διαστήματος $[t_1 \ t_2]$ (Αν επιλέξουμε τις συνιστώσες που έχουν μέγιστη διακύμανση εντος του $[t_1 \ t_2]$ είναι το ίδιο.).

4.2.1.2 Πίνακας Μείξης

Απορρίπτονται οι συνιστώσες των οποίων ο πίνακας μείξης ικανοποιεί τις παρακάτω προϋποθέσεις:

1. Ο πίνακας μείξης δεν είναι τετράγωνικός
2. Ο πίνακας μείξης δεν έχει τις ίδιες διαστάσεις με τον αριθμό των ηλεκτροδίων

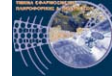
4.2.1.3 Zero-crossings

Οι παρασιτικές συνιστώσες που εντοπίζονται και απορρίπτονται με βάση πόσες φορές περνούν τον άξονα του μηδενός(zero-crossings). Αν λοιπόν ο αριθμός των zero-crossings μιας συνιστώσας ξεπερνά το όριο του [μέσος + τυπική απόκλιση] των zero-crossings για όλες τις συνιστώσες, τότε απορρίπτεται ως παρασιτική.

Πρέπει να επισημάνθει ότι, αυτή η μέθοδος δεν απορρίπτει όλες τις παρασιτικές συνιστώσες, όμως αποτελεί ένα εύκολο τρόπο απόρριψης σε αντίθεση με τον εξονυχιστικό οπτικό έλεγχο των συνιστωσών.

4.2.2 Επιλογή Χαρακτηριστικών για Ομαδοποίηση

Από τις συνιστώσες που έχουν περάσει από τους παραπάνω ελέγχους, εξάγουμε τα εξής χαρακτηριστικά:



- Απόλυτες τιμές για θετικές και αρνητικές κορυφές των συνιστωσών εντός του $[t_1 \ t_2]$.
- Τη χρονική στιγμή κατά την οποία συμβαίνει η κορυφή
- Τις αντίστοιχες στήλες από τους πίνακες μείξης

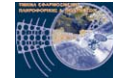
4.2.3 Διαδικασία Ομαδοποίησης

Η διαδικασία ομαδοποίησης περιλαμβάνει δύο στάδια:

1^η ομαδοποίηση: οι συνιστώσες ομαδοποιούνται με βάση το πλάτος(απόλυτες τιμές) και τη χρονική στιγμή των κορυφών (χρονική ομαδοποίηση).

2^η ομαδοποίηση: συνιστώσες ομαδοποιούνται με βάση τις αντίστοιχες στήλες του πίνακα μείξης ή τις θέσεις των διόλων των αντίστοιχων πηγών (τοπογραφική ομαδοποίηση).

Με αυτά τα δύο στάδια, πετυχαίνουμε χρονική και τοπογραφική ομαδοποίηση.

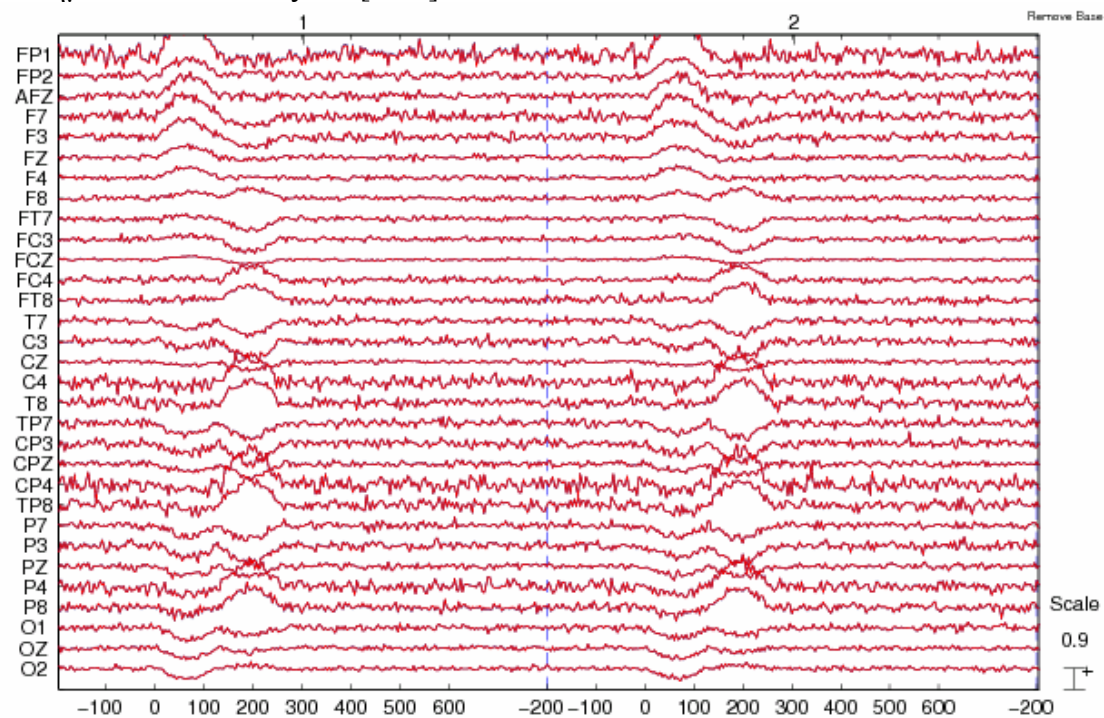


4.3 Περιγραφή αποτελεσμάτων

4.3.1 Διαδικασία προεπεξεργασίας δεδομένων

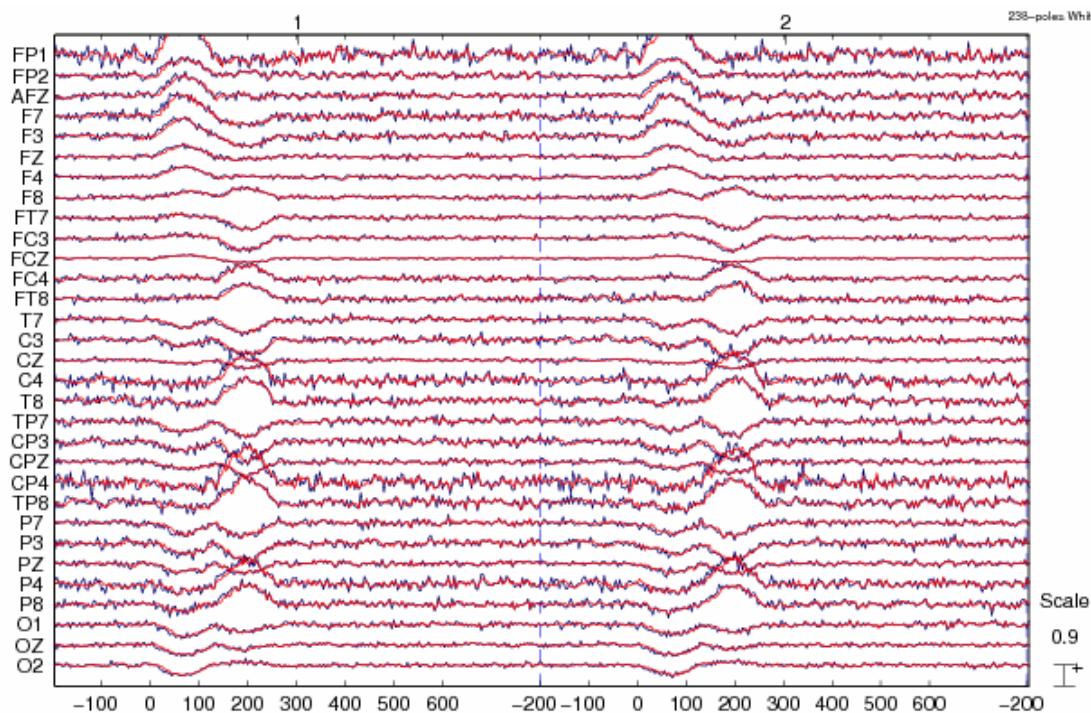
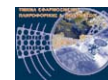
Πριν την εφαρμογή της ICA ακολουθείται η διαδικασία κεντραρίσματος των δεδομένων στο μηδέν και η εφαρμογή ενός φίλτρου λεύκανσης.

Το κεντράρισμα πραγματοποιείται με αφαίρεση του μέσου που έχει υπολογιστεί για τα σημεία-frames εκτός του $[t1 \ t2]$.



Σχ.16 Κεντράρισμα των δεδομένων με αφαίρεση του μέσου(υπολογισμένος για τις εκτός του $[t1 \ t2]$). Τα αρχικά σήματα σχεδιάζονται με μπλε και τα νεότερα με κόκκινο.

Το φίλτρο λεύκανσης μοντελοποιεί τις χρονοσειρές υπολογίζοντας συντελεστές γραμμικής πρόβλεψης. Ο βαθμός του μοντέλου που εφαρμόζεται είναι ίσος με το 95% των συνολικών σημείων-frames των χρονοσειρών.

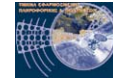


Σχ.17 Ένα p -βαθμού LPC φίλτρο, όπου p είναι το 95% των σημείων της χρονοσειράς. Τα αρχικά σήματα σχεδιάζονται με μπλε και τα νεότερα με κόκκινο.

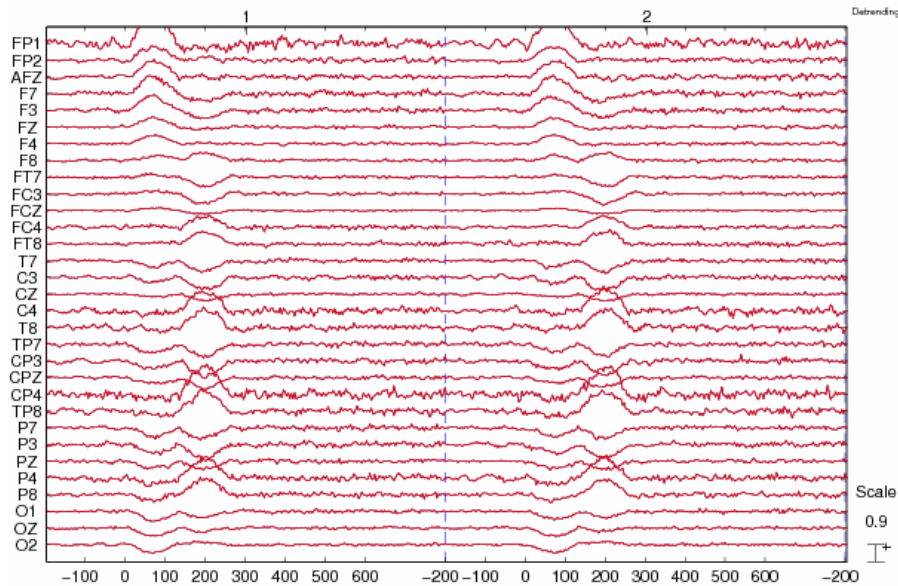
4.3.2 Ανάλυση και οπτικοποίηση πριν την ICA

Επίσης εφαρμόζονται κάποιες επιπλέον μέθοδοι ελέγχου για τα δεδομένα, οι οποίες έχουν αναπτυχθεί με σκοπό την απόρριψη ή απλά τον εντοπισμό ιδιαίτερα παρασιτικών ηλεκτροδίων ή trials. Οι χρονοσειρές που τριγγάρουν στους εξής ελέγχους:

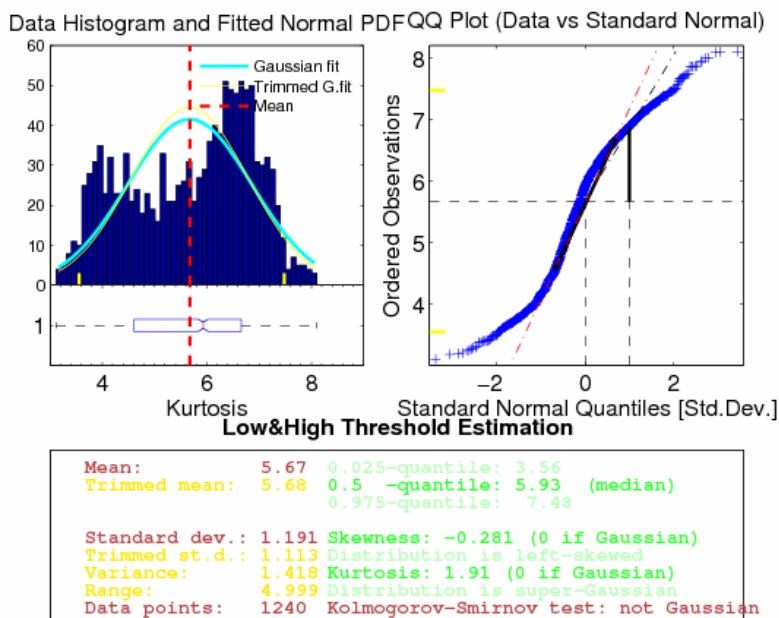
- Έλεγχος για τιμές που κόβονται στο όριο καταγραφής τάσεων, δηλ. αν για θετικές και αρνητικές τάσεις δίνεται εύρος καταγραφής, από τον εγκεφαλογράφο, 16 bit, τότε οι τιμές που καταγράφονται δεν μπορούν να υπερβαίνουν τα 8 bit.
- Φασματικός έλεγχος για μυϊκά και οπτικά παράσιτα [42]
- Έλεγχος και διόρθωση για πιθανή ύπαρξη γραμμικής κλίσης(20°) σε κάθε χρονοσειρά [42]



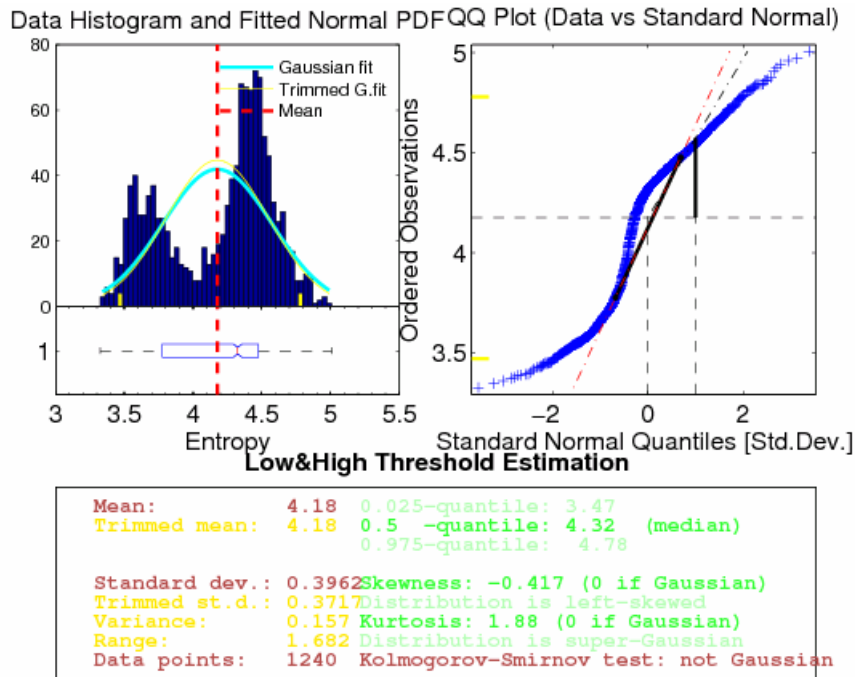
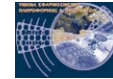
- Έλεγχος των στατιστικών μέτρων της κύρτωσης, της εντροπίας και της κοινής πιθανότητας. Τα μέτρα αυτά σχετίζονται άμεσα με τη ικανότητα διαχωρισμού της ICA όπως αναπτύχθηκε άνωθεν.[42]



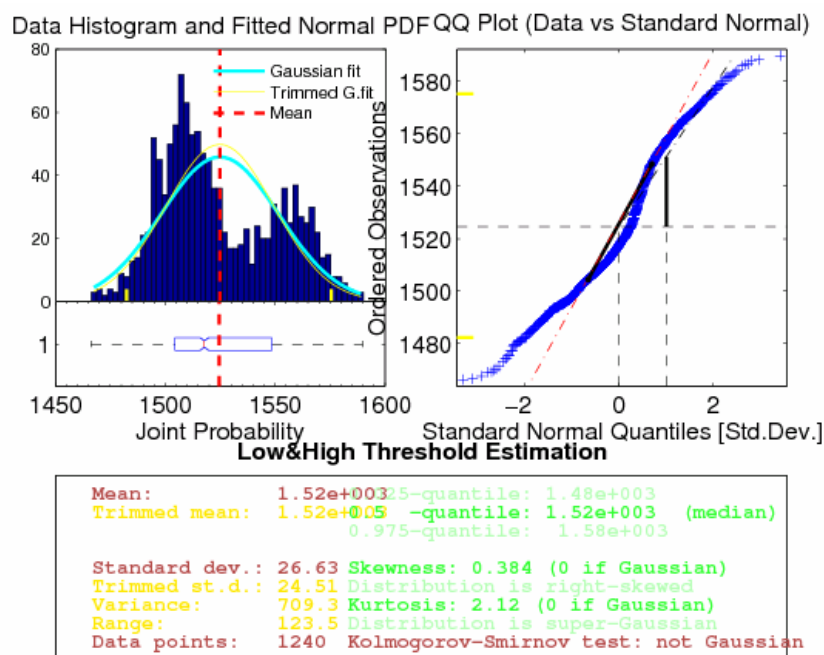
Σχ.18 Ένα δείγμα ελέγχου για γραμμικές κλίσεις(20°), όπως παρατηρείται στα δείγματα δεν έχει εντοπιστεί κάποια κλίση ώστε να διορθωθεί. Τα αρχικά σήματα σχεδιάζονται με μπλε και τα νεότερα με κόκκινο.



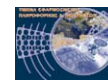
Σχ.19 Ένα δείγμα στατιστικής ανάλυσης σχετικά με την κύρτωση που παρουσιάζουν τα κανάλια του μπλοκ των δεδομένων.



Σχ.20 Ένα δείγμα στατιστικής ανάλυσης σχετικά με την εντροπία που παρουσιάζουν τα κανάλια του μπλοκ των δεδομένων.

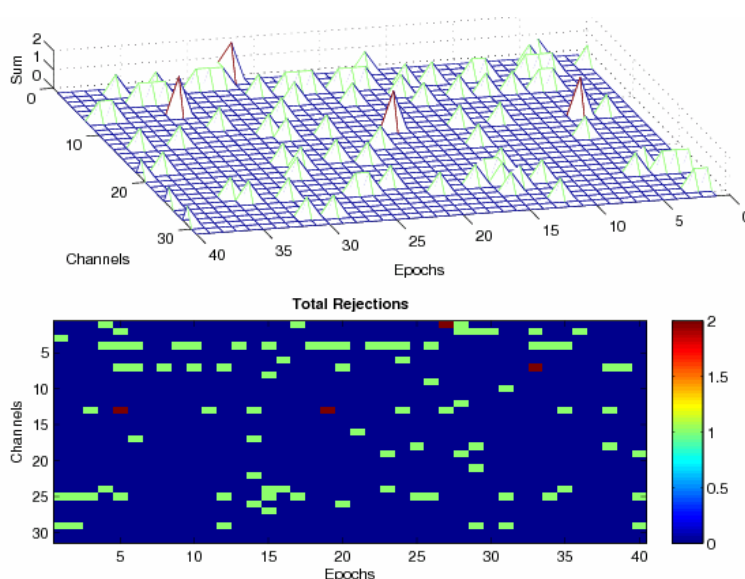


Σχ.21 Ένα δείγμα στατιστικής ανάλυσης σχετικά με την κοινή πληροφορία που παρουσιάζουν τα κανάλια του μπλοκ των δεδομένων.



Στη συνέχεια ακολουθεί οπτικοποίηση των αποτελεσμάτων της παραπάνω μεθόδου ελέγχου με ένα γράφημα που παρουσιάζει μία κάτοψη ηλεκτροδίων και trial. Κάθε ηλεκτρόδιο για κάθε φορά που τριγγάρεται ως παρασιτικό παίρνει μία μονάδα, έτσι ανάλογα με τα βήματα που τριγγάρει παρουσιάζει και το ανάλογο άθροισμα στο γράφημα.

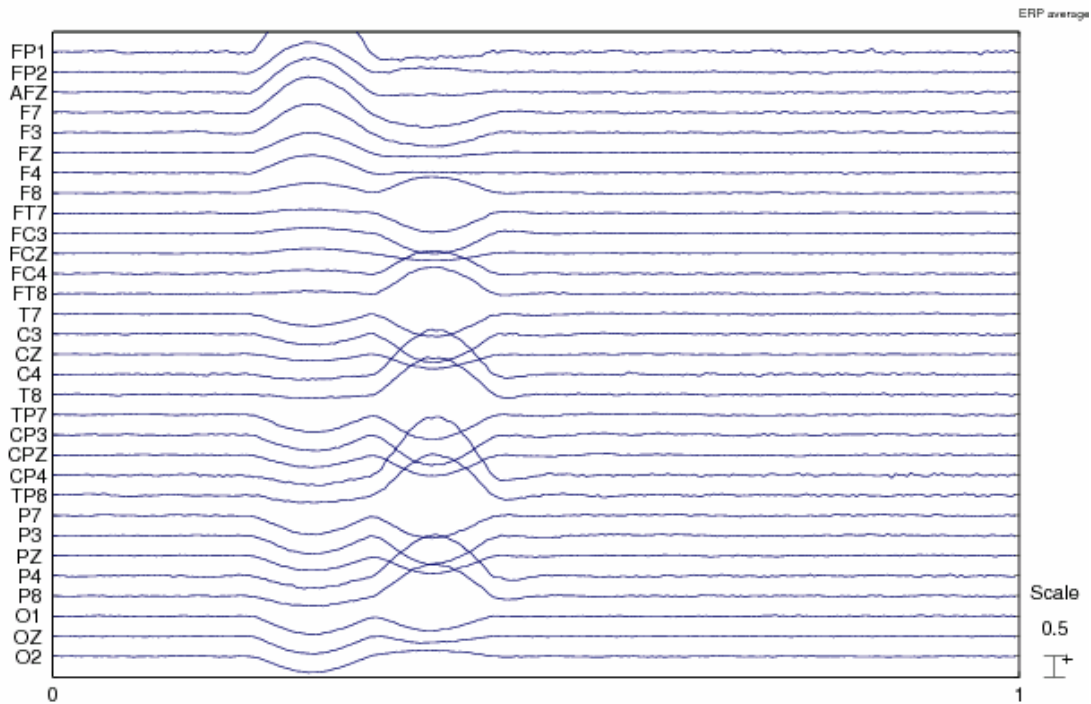
Αυτή η μέθοδος ελέγχου χρησιμοποιείται μόνο ως μία αρχική οπτικοποίηση της καταλληλότητας των δεδομένων να επεξεργαστούν από την ICA, η βασική μεθοδολογία (κεφ. 4.2) που ακολουθήται δεν μας επιτρέπει να απορρίψουμε κανένα ηλεκτρόδιο ή trial.



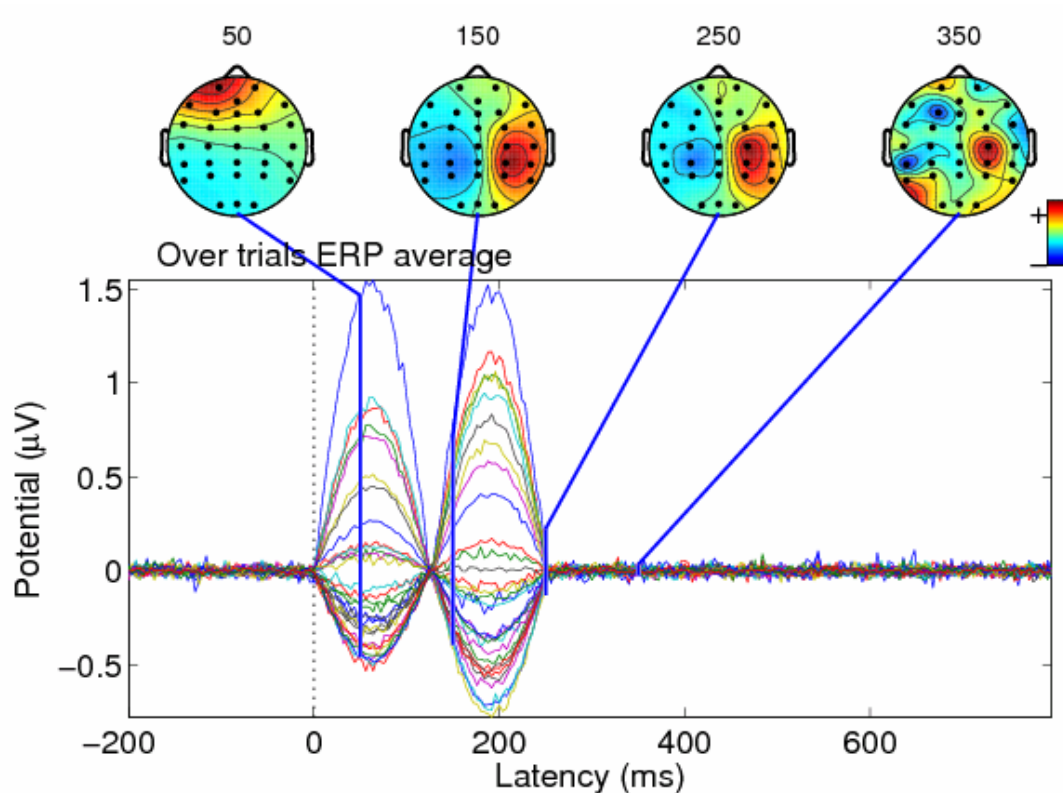
Σχ.22 Όπως φαίνεται στο σχήμα, η οπτικοποίηση βοηθάει για να αντιληφθούμε αν υπάρχουν προβληματικά κανάλια ή trials.

4.3.3 Οπτικοποίηση του μέσου EP

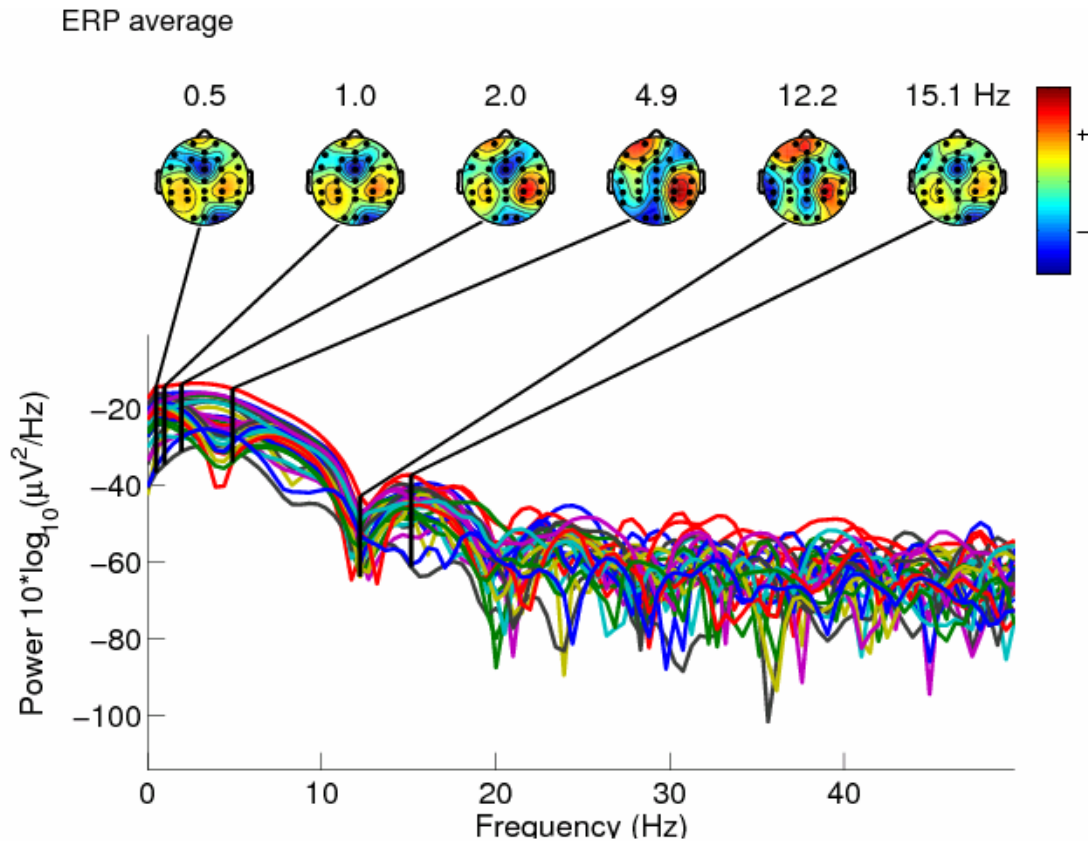
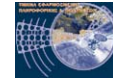
Όπως αναφέρθηκε άνωθεν (κεφ. 2) ο υπολογισμός του μέσου για όλα τα trials εφαρμόζεται ώστε να παραμείνει η βασική πληροφορία που χαρακτηρίζει το EP. Με μόνο σκοπό να αντιληφθούμε το σχήμα του EP οπτικοποιούνται τα αποτελέσματα του μέσου όπως φαίνεται ακολούθως στα σχήματα.



Σχ.23 Στο σχήμα φαίνεται το μέσο EP για όλα τα trials για κάθε ηλεκτρόδιο.



Σχ.24 Οι τοπογραφίες που παρουσιάζονται δείχνουν την κρανιακή ένταση για τις χρονικές στιγμές των εκάστοτε κορυφών.

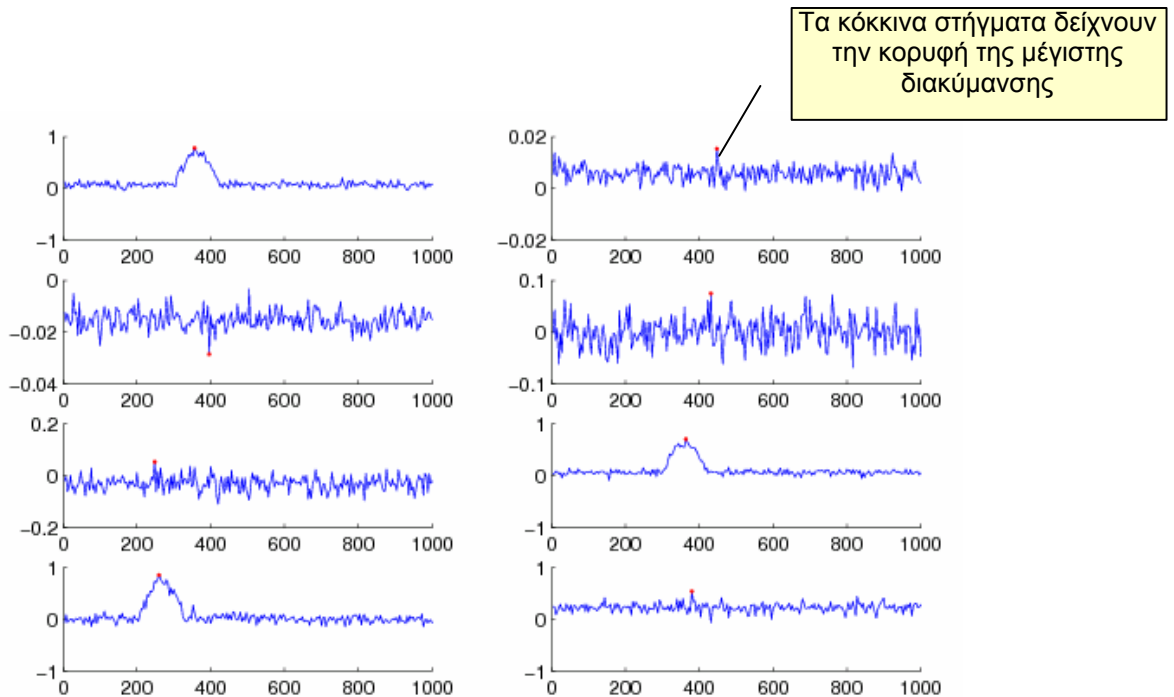


Σχ.25 Οι τοπογραφίες που παρουσιάζονται δείχνουν την φασματική ένταση από τοπογραφική οπτική που αφορούν κάποιες σημαντικές συχνότητες του μέσου EP. Οι συχνότητες παρουσιάζονται στη συγκεκριμένη μπάνα γιατί μετά τα 20Hz η φασματική ισχύς μειώνεται με βήμα άνω των 40dB.

4.3.4 Εντοπισμός της μέγιστης διακύμανσης

Ο εντοπισμός της μέγιστης διακύμανσης αποτελεί ένα πολύ σημαντικό σημείο της μελέτης των δεδομένων. Επιβάλλεται να είναι ιδιαίτερα αποτελεσματική, για το γεγονός ότι η 1^η ομαδοποίηση θα βασιστεί στα εξαγόμενα χαρακτηριστικά του συγκεκριμένου βήματος. Σύμφωνα με την περιγραφή της διακύμανσης ως ενέργεια από τον Hyvärinen[43], ο εντοπισμός επικεντρώνεται στις κορυφές που παρουσιάζουν μέγιστη ενέργεια.

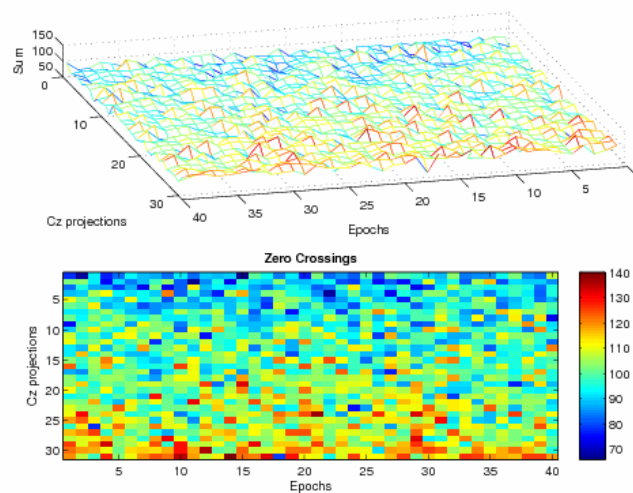
Παρακάτω παρουσιάζονται κάποια τυχαία δείγματα, από τις επεξεργαζόμενες προβολές στο Cz, στα οποία έχει εντοπιστεί η χρονική στιγμή της μέγιστης διακύμανσης.



Σχ.26 Εντοπισμός διακύμανσης στα επεξεργαζόμενα σήματα.

4.3.5 Απόρριψη zero-crossings

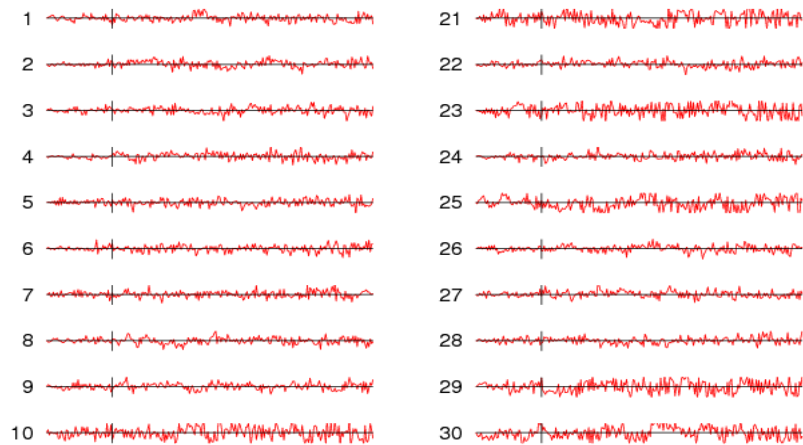
Ακολουθώντας τα βήματα της μεθόδου (κεφ 4.2.1.3) έχει υπολογιστεί για κάθε Cz προβολή ο αριθμός των zero-crossings και σύμφωνα με το όριο απόρριψης έχουν αποκλειστεί από την ομαδοποίηση προβολές όπως φαίνεται παρακάτω.



Σχ.28 Στο σχήμα παρουσιάζονται οι τιμές των zero-crossings για κάθε Cz προβολή.



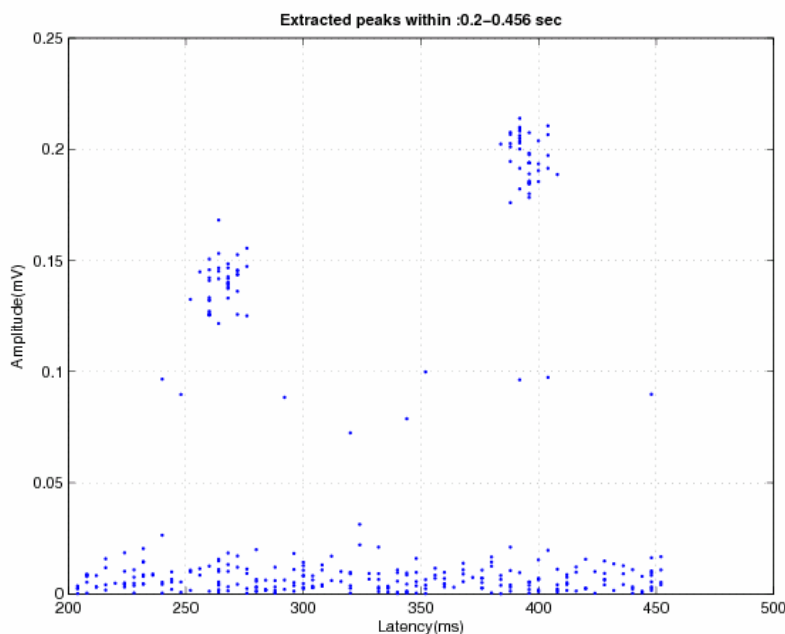
Zero crossing rejected samples



Σχ.29 Τυχαία επιλεγμένες Cz προβολές που έχουν απορριφθεί από τη μέθοδο.

4.3.6 Εξαγωγή χαρακτηριστικών για τα δύο στάδια ομαδοποίησης

Σε αυτό το σημείο για όλες τις προβολές που πέρασαν από τις διαδικασίες επιλογής και απόρριψης ακολουθεί η εξαγωγή των χαρακτηριστικών για κάθε μία από τις προβολές.



Σχ.30 Τα χαρακτηριστικά όλων των κορυφών μετά τη διαδικασία του zero-crossing.

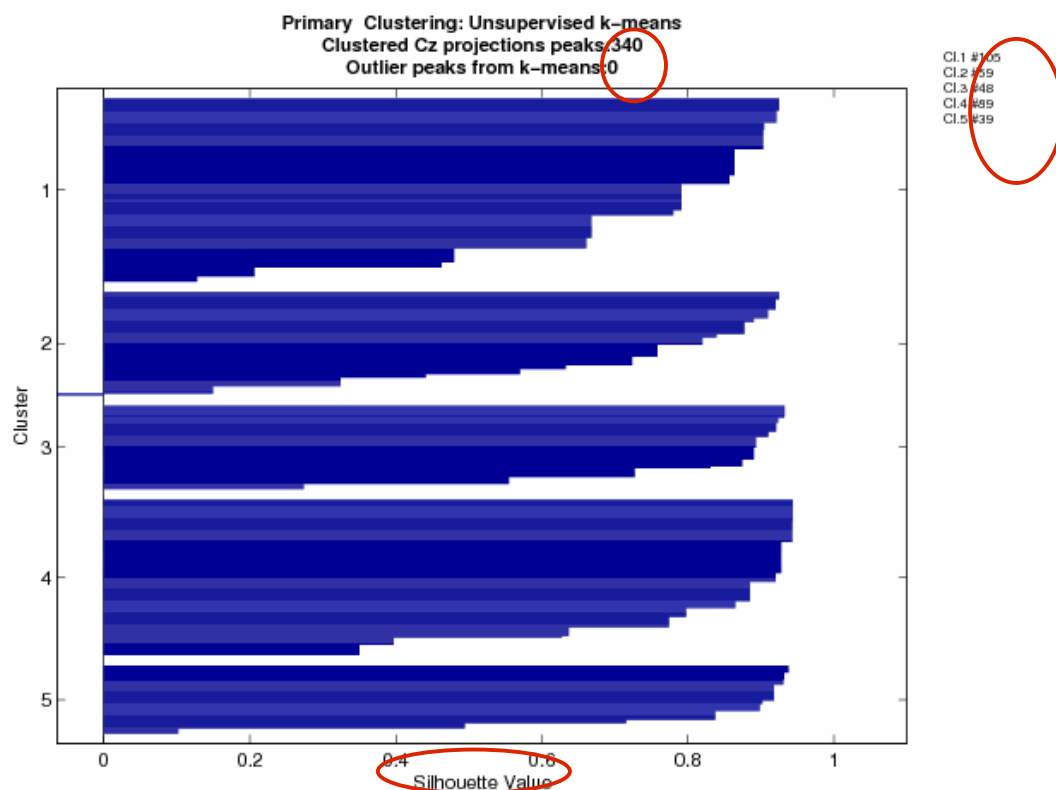


4.3.7 1^η ομαδοποίηση – Χαρακτηριστικά των κορυφών με τη μέγιστη διακύμανση

Τα χαρακτηριστικά που εξάγονται από το προηγούμενο βήμα επεξεργασίας ομαδοποιούνται με τη χρήση του αλγόριθμου των κ-μέσων. Οι αποστάσεις των δειγμάτων από τα κέντρα των ομάδων, τα οποία έχουν επιλεγεί αρχικά σε τυχαίες θέσεις, έχουν υπολογιστεί για την τετραγωνική ευκλίδεια απόσταση.

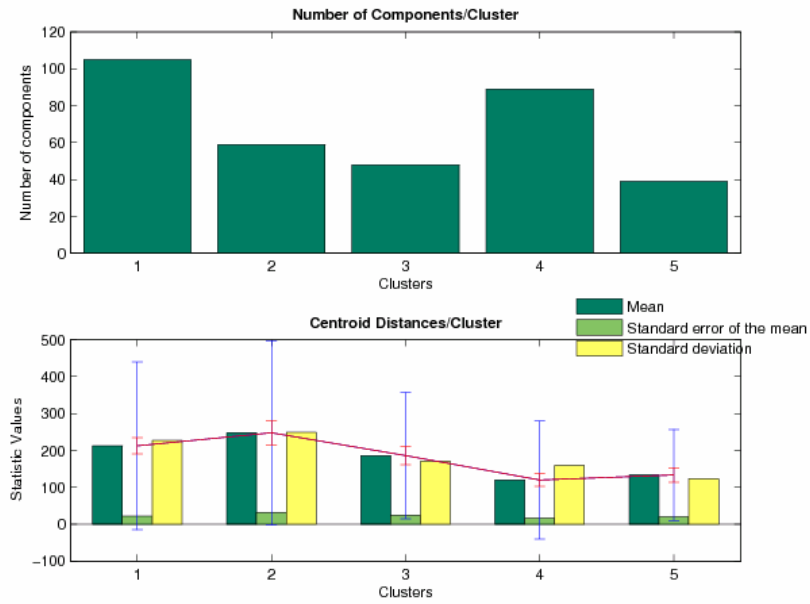
Οι αρχικές ομάδες αν περιέχουν δείγματα που απέχουν απόσταση πάνω από την 5πλάσια της τυπικής απόκλισης των υπολοίπων δειγμάτων της ομάδας εξαιρούνται. Η διαδικασία επαναλαμβάνεται για 10 επαναλήψεις ή έως ότου έχουν δημιουργηθεί συμπαγείς ομάδες πριν την ολοκλήρωση των επαναλήψεων.

Τα 340 δείγματα θα καταταχθούν σε 5 ομάδες, από τα οποία μόνο και τα 340 ομαδοποιούνται όπως φαίνεται παρακάτω.

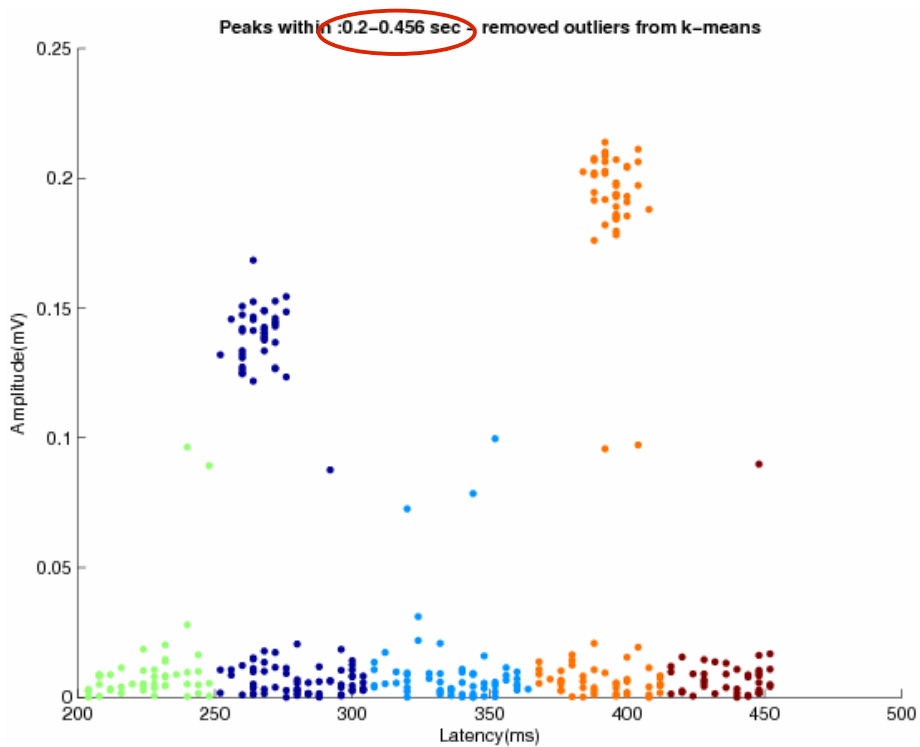


Σχ.32 Αυτό το γράφημα παρουσιάζει την τιμή silhouette κάθε δείγματος.

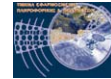
Επίσης τυπώνονται κάποιες λεπτομέρειες για το πλήθος των δειγμάτων.



Σχ.33 Παρουσιάζονται τα ιστογράμματα του πληθυσμού των ομάδων καθώς και απλές στατιστικές τιμές για τις αποστάσεις των δειγμάτων από τα κέντρα των ομάδων.



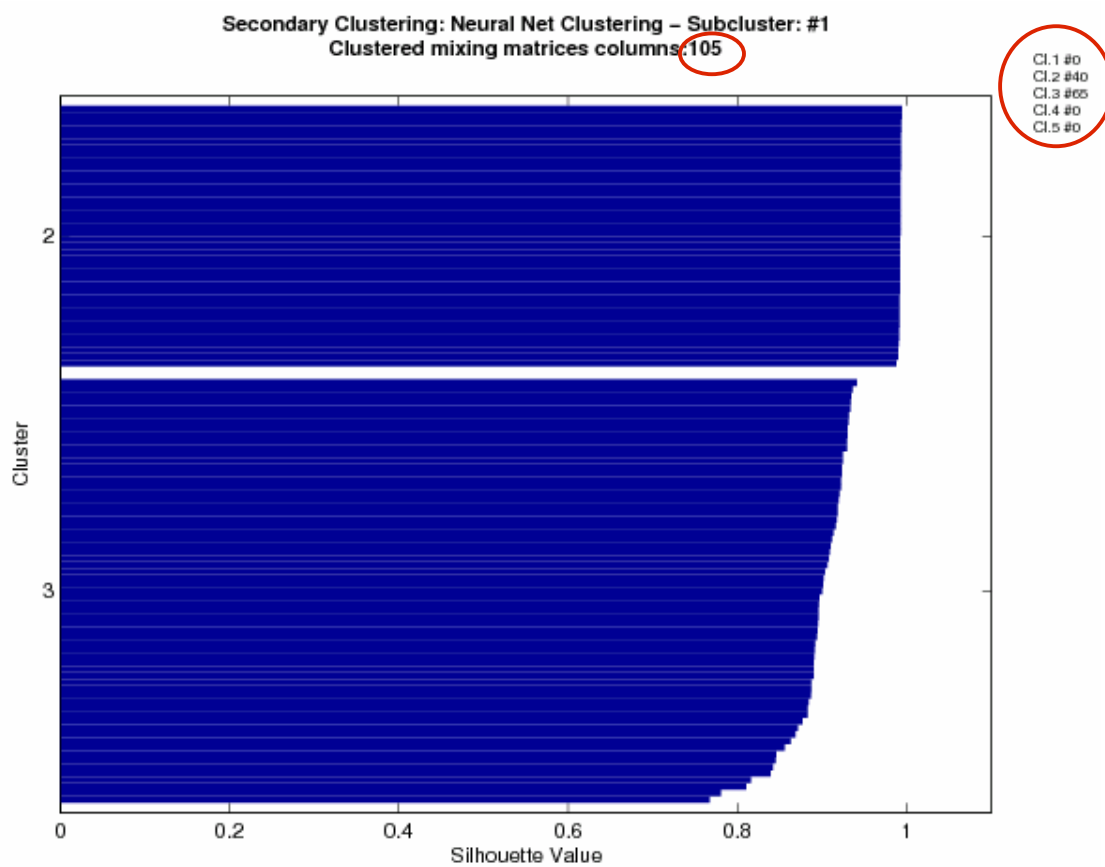
Σχ.34 Το παραπάνω διάγραμμα διασποράς (scatter plot) δείχνει ομαδοποιημένα δείγματα.



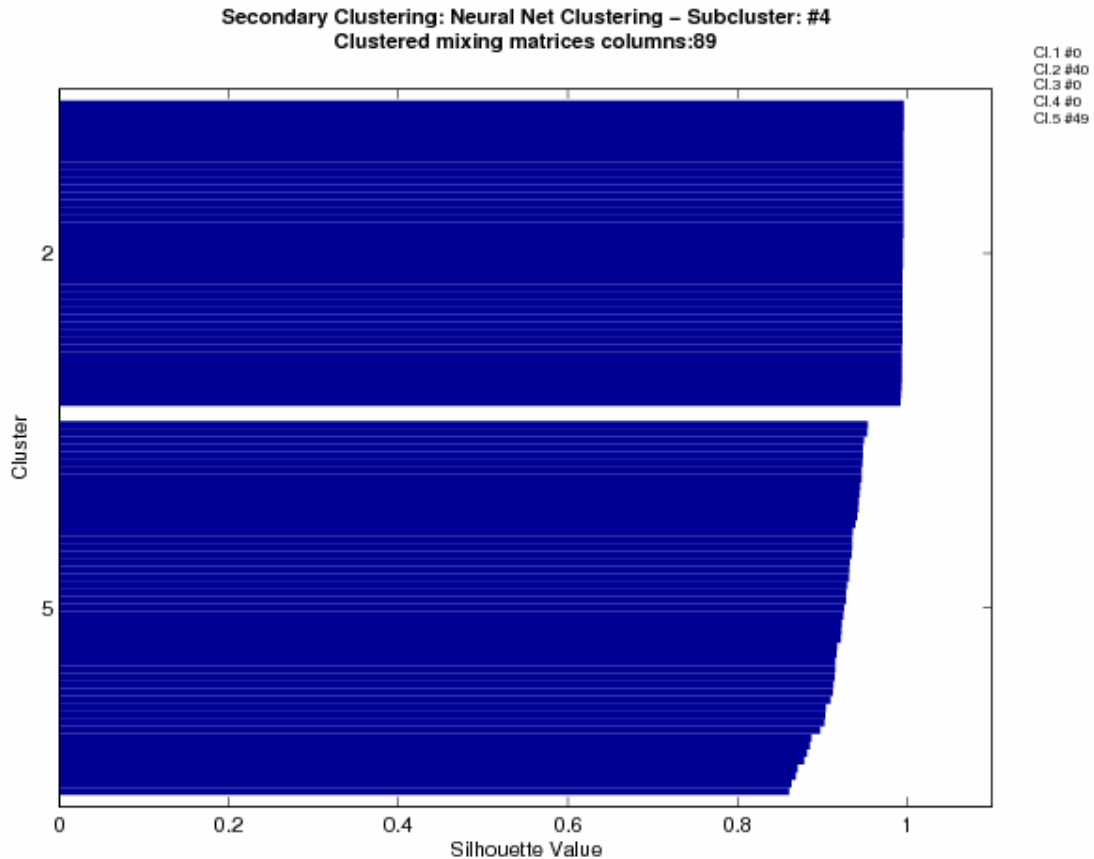
4.3.8 2^η ομαδοποίηση - Στήλες των πινάκων μείξης

Αμέσως μετά την 1^η ομαδοποίηση για τις 5 ομάδες που έχουν δημιουργηθεί ακολουθεί το 2^ο στάδιο για την εξόρυξη των EPs. Η ομαδοποίηση επιτυγχάνεται με τη χρήση ενός νευρωνικού δικτύου που αποτελείται από 5 νευρώνες.

Το παρακάτω σχήμα παραθέτει στοιχεία για το πλήθος και τις τιμές silhouette των ομαδοποιημένων.



Σχ.36 Το γράφημα 2^{ης} ομαδοποίησης την 1^η ομάδα. Η υπό-ομάδα 2 αποτελείται από εξαιρετικά όμοια δείγματα. Όπως θα φανεί στη συνέχεια περιέχει τα ομαδοποιημένα EPs. Το διάγραμμα των silhouette values αγγίζει το 100% της ομοιομορφίας των δειγμάτων και σχεδόν έχει ορθογώνιο σχήμα.

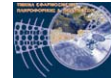


Σχ.37 Το γράφημα 2^{ης} ομαδοποίησης για την 4^η ομάδα. Η υπό-ομάδα 2 αποτελείται από εξαιρετικά όμοια δείγματα. Όπως θα φανεί στη συνέχεια περιέχει τα ομαδοποιημένα EPs. Το διάγραμμα των silhouette values αγγίζει το 100% της ομοιομορφίας των δειγμάτων και σχεδόν έχει ορθογώνιο σχήμα.

4.3.8.1 Οπτικοποίηση τελικών υπο-ομάδων

Έπειτα από τη 2^η ομαδοποίηση σύμφωνα με τα βήματα που έχουν αναληθεί στο κεφάλαιο 4.2 παρουσιάζεται ένα δείγμα οπτικοποίησης των τελικών ομάδων.

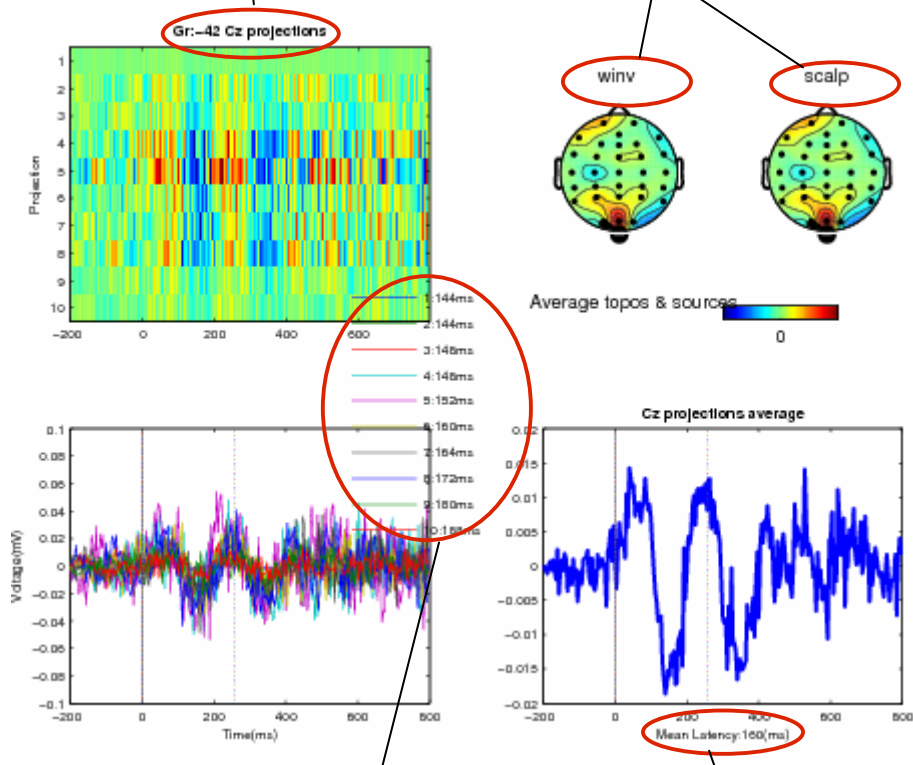
Σε όλα τα γραφήματα ισχύει η ταξινόμηση με βάση τη χρονική εμφάνιση της μέγιστης κορυφής για κάθε προβολή.



Επισημανση στις σημαντικές λεπτομέρειες που αναγράφονται στα γραφήματα

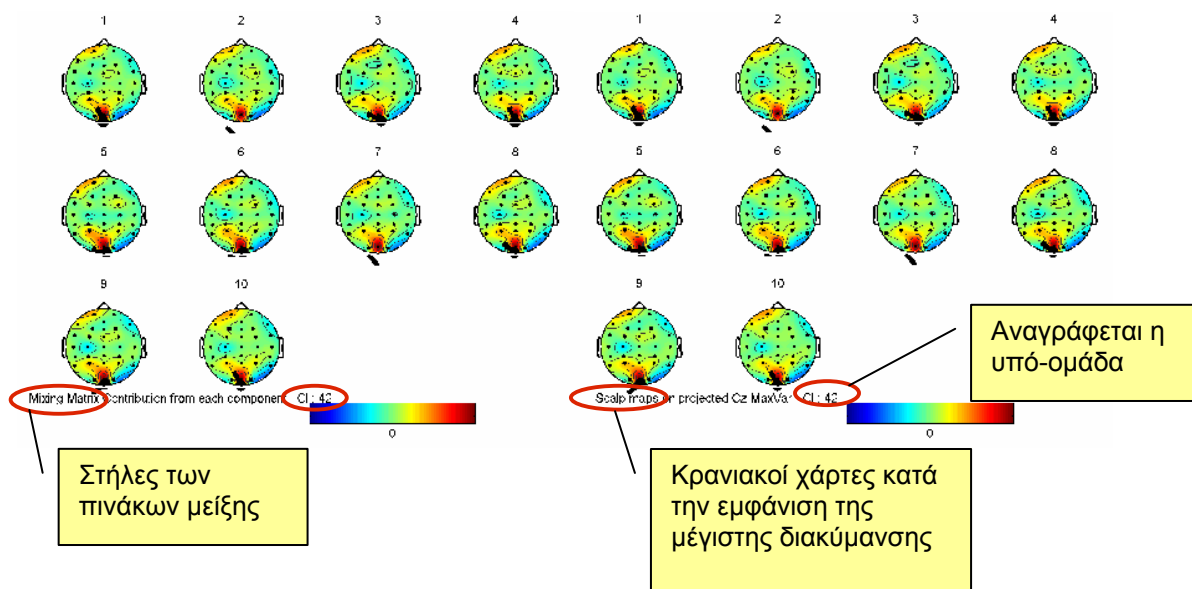
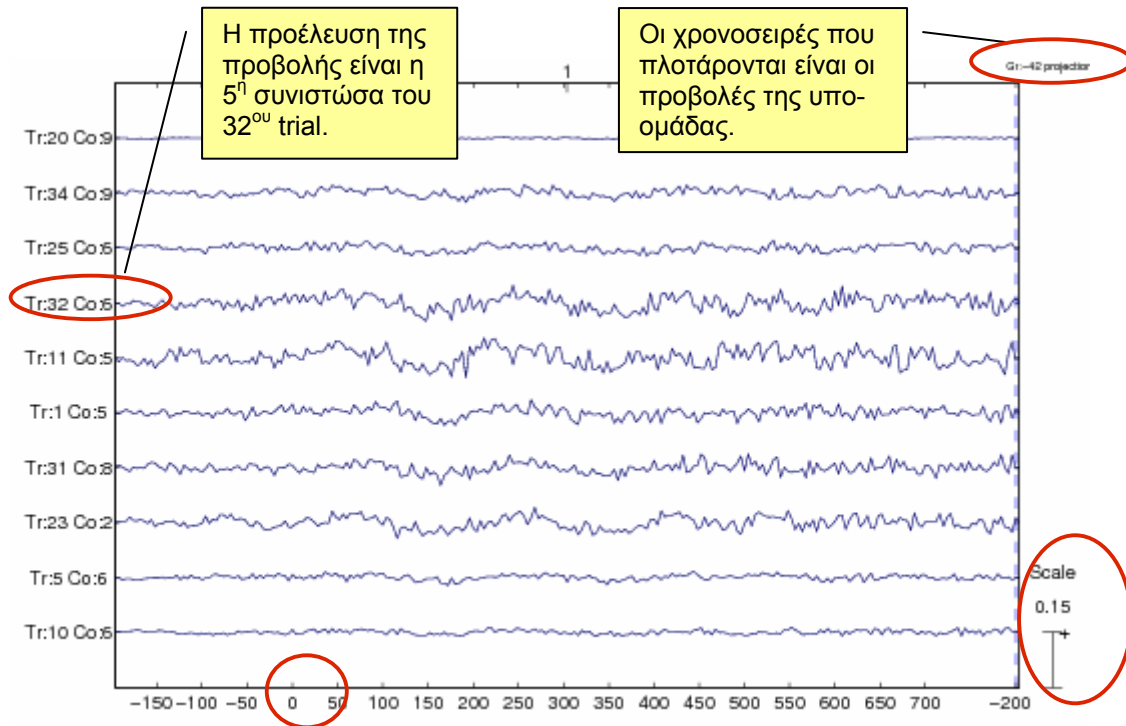
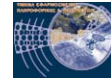
Cz προβολές της ομάδας
Gr:-42 όπου,
 “-”: υπό-ομάδα αρνητικών κορυφών
 “4”: 4^η ομάδα της 1^{ης} ομαδοποίησης
 “2”: 2^η υπο-ομάδα της 4^{ης} ομάδας

“winv”: η μέση στήση των πινάκων μείξης
 “scalp”: ο μέσος κρανιακός χάρτης κατά τη χρονική εμφάνιση της μέγιστης κορυφής



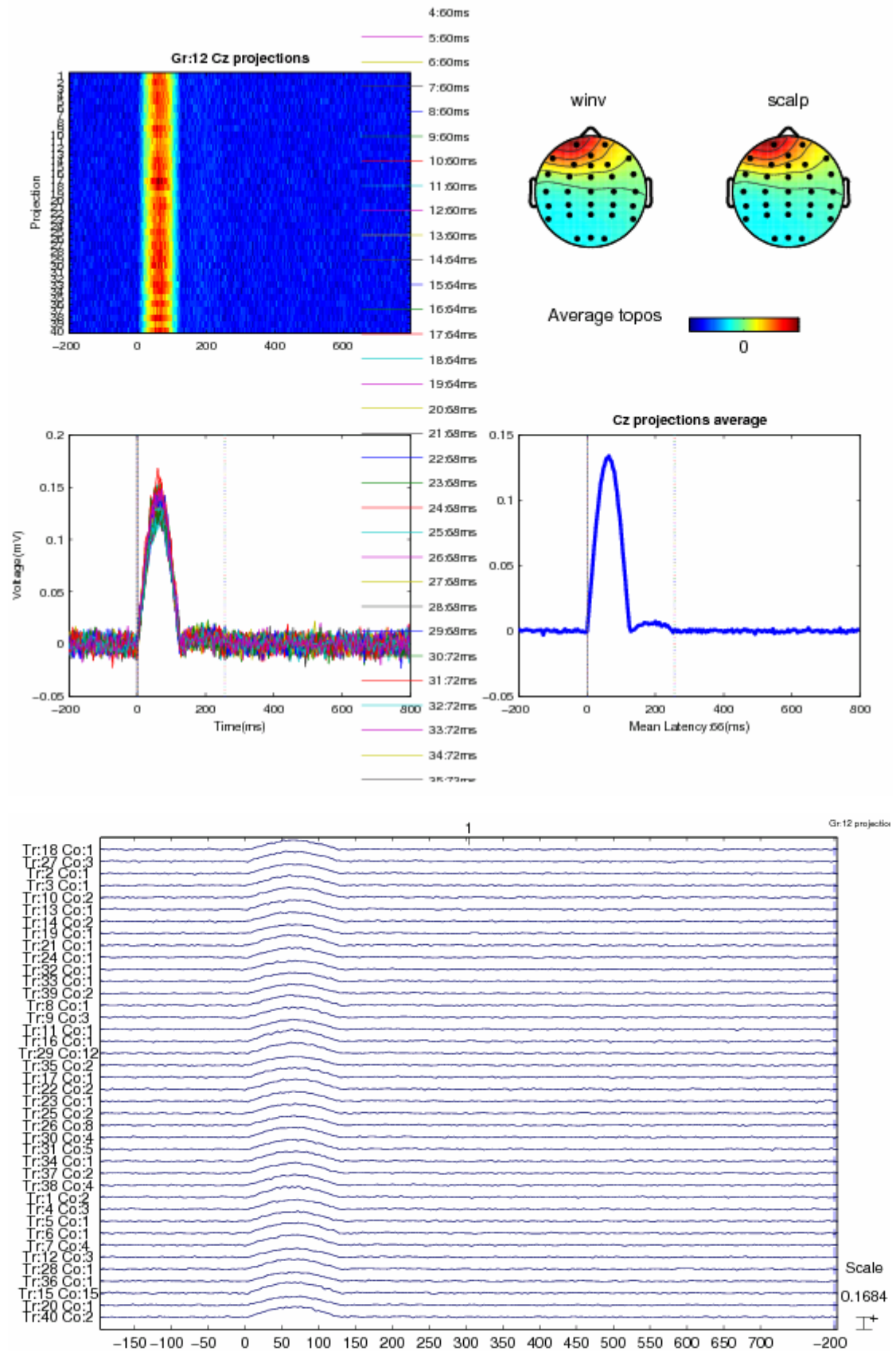
Οι προβολές ταξινομούνται βάση της χρονικής εμφάνισης της μέγιστης κορυφής.

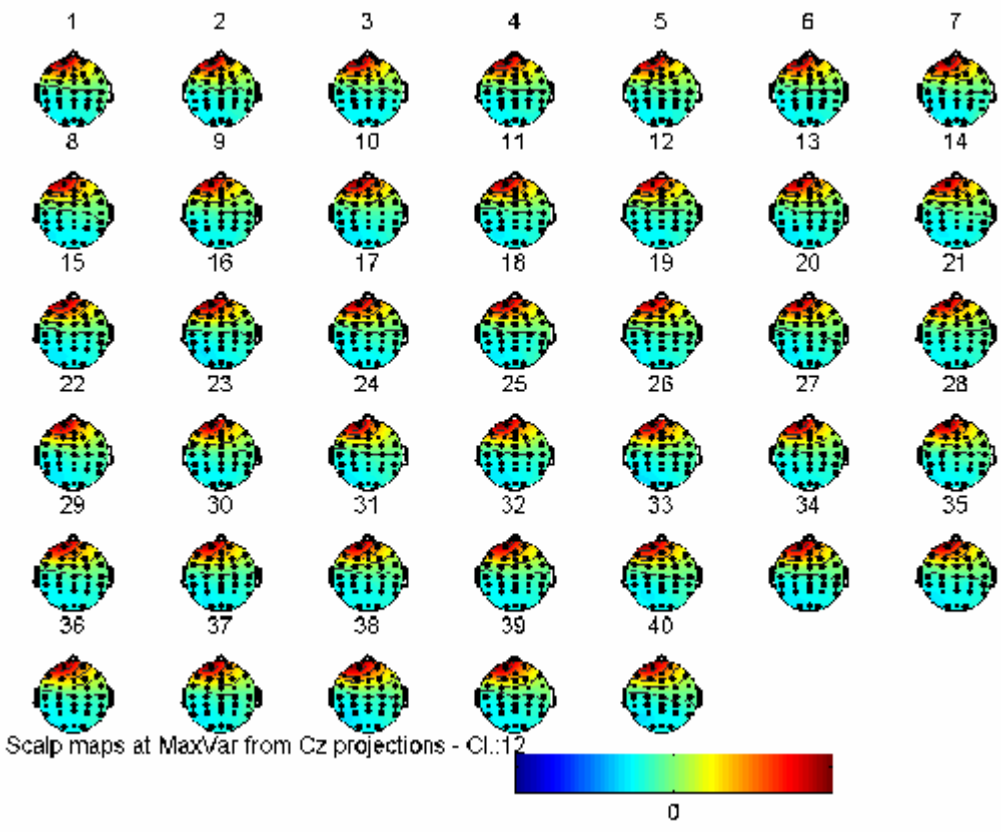
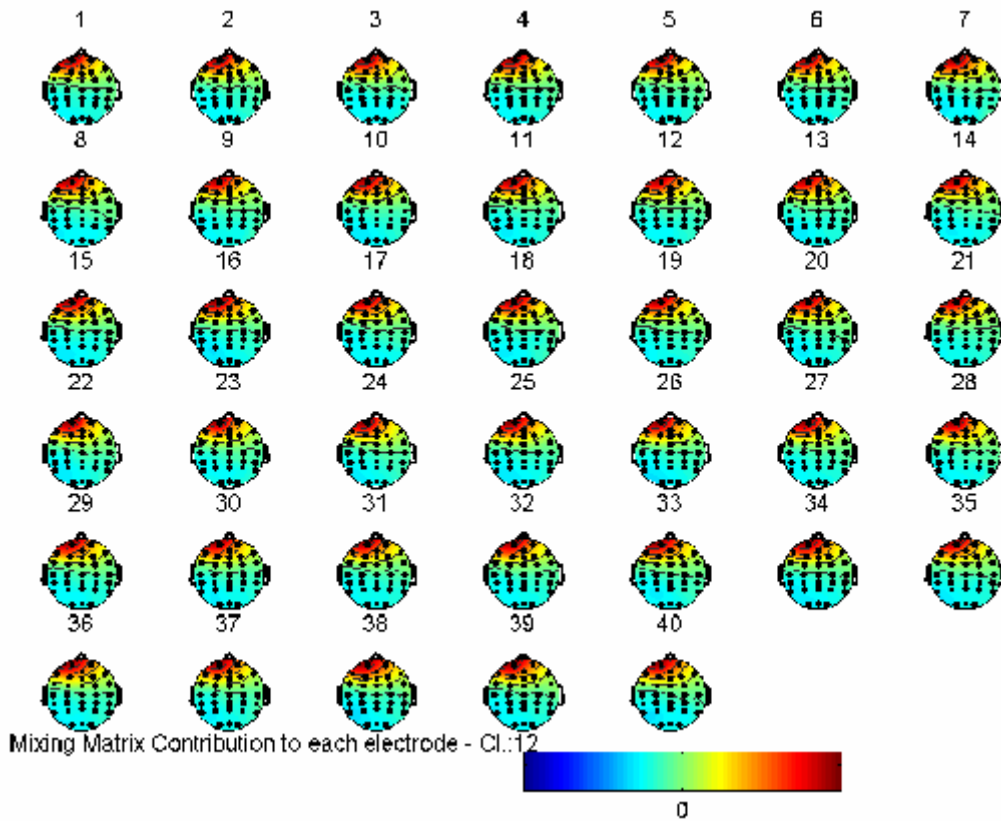
Η μέση χρονική στιγμή της υπό-ομάδας.

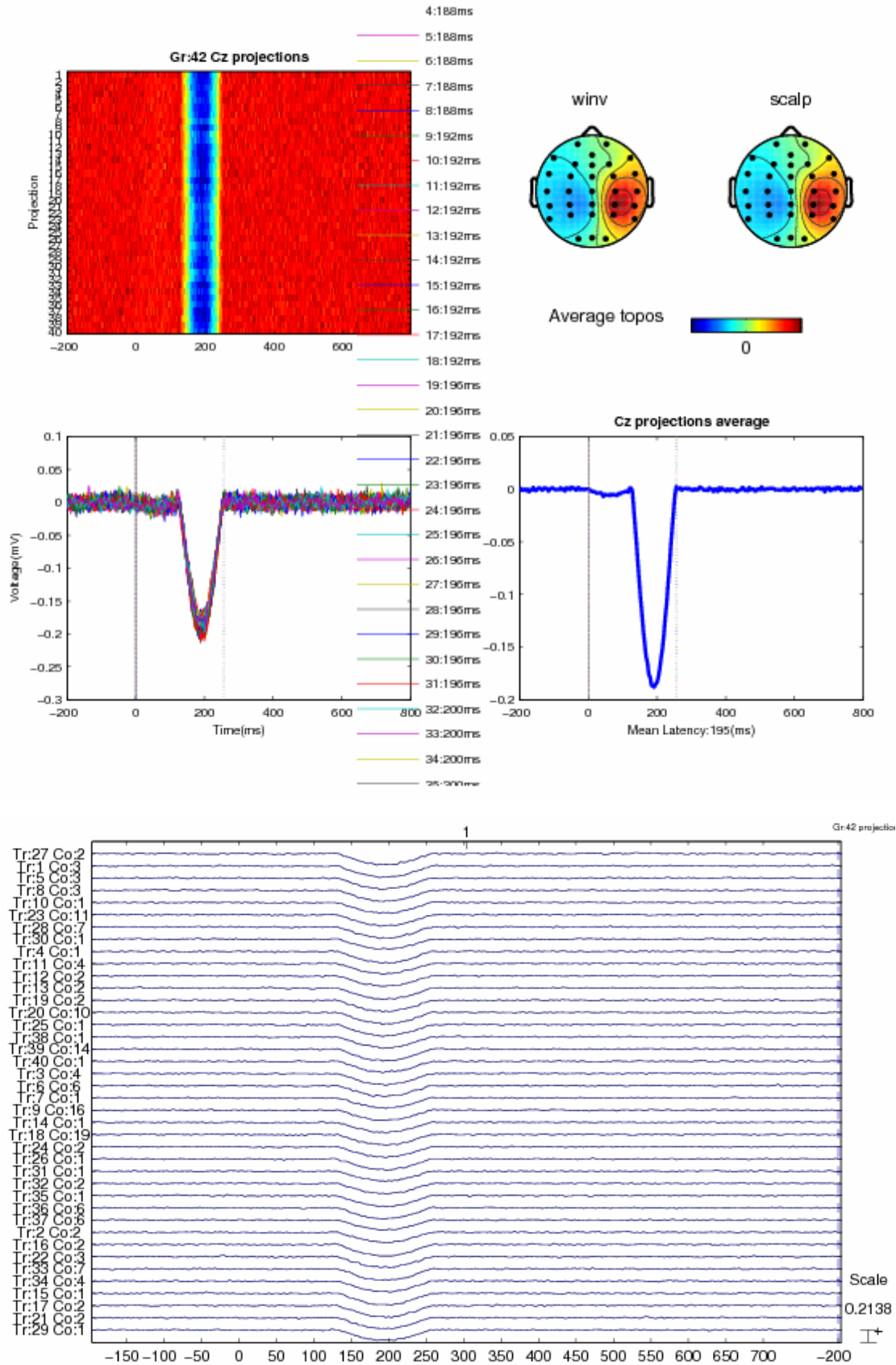
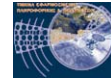


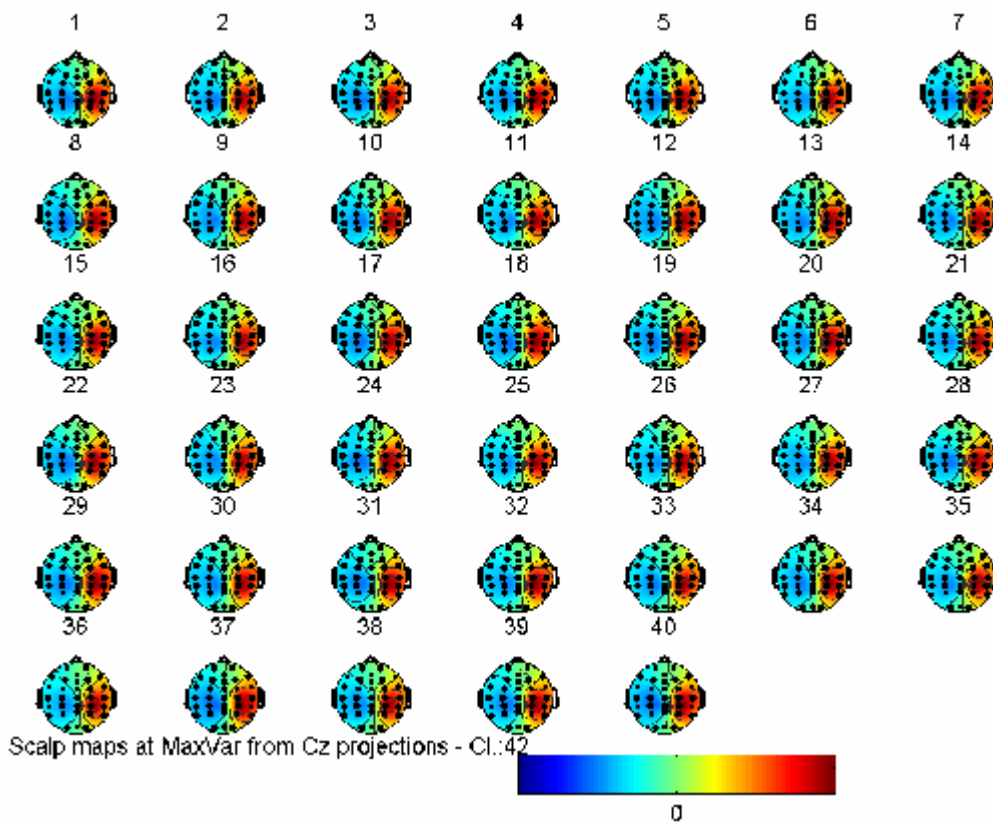
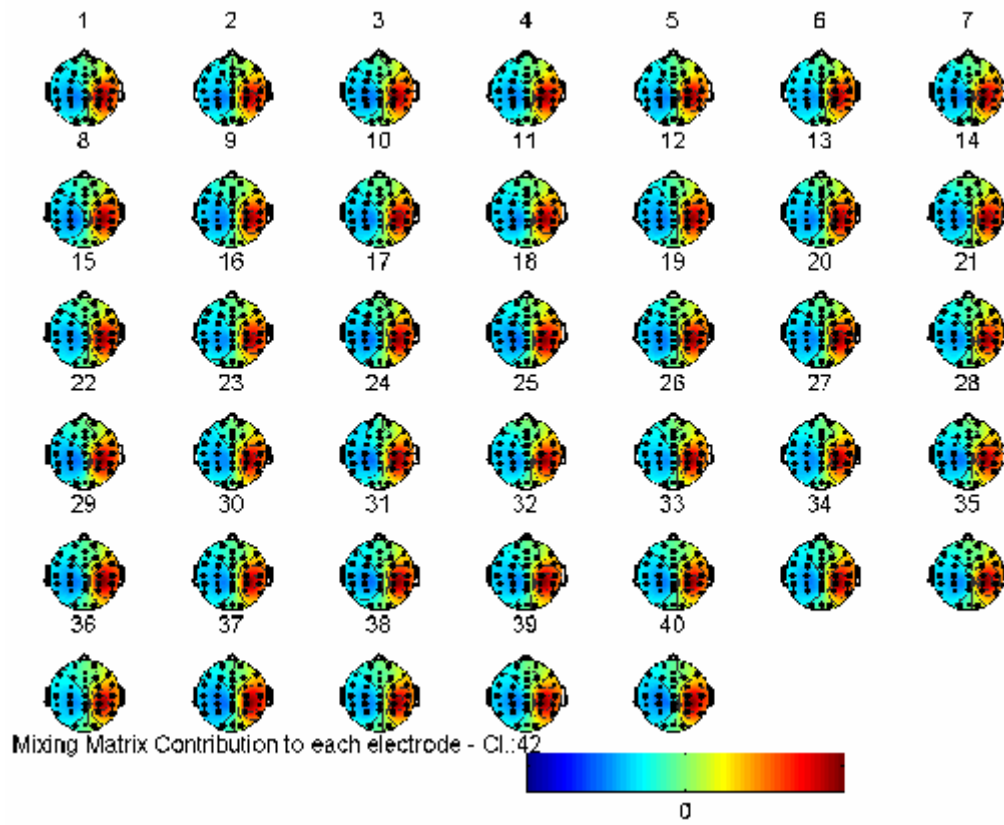


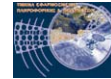
4.3.8.2 Οι υπο-ομάδες που αποτελούνται από τις αρχικές πηγές



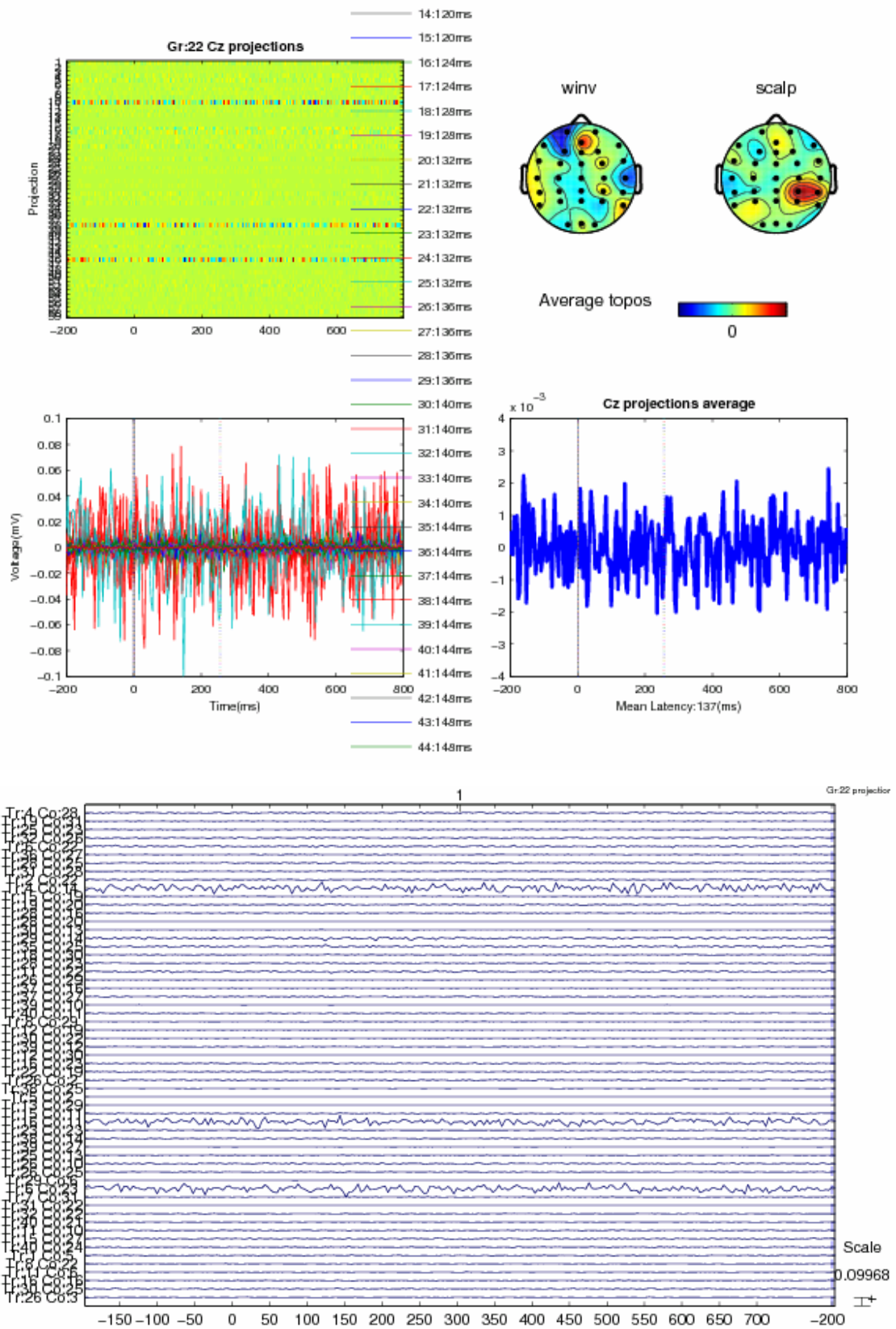


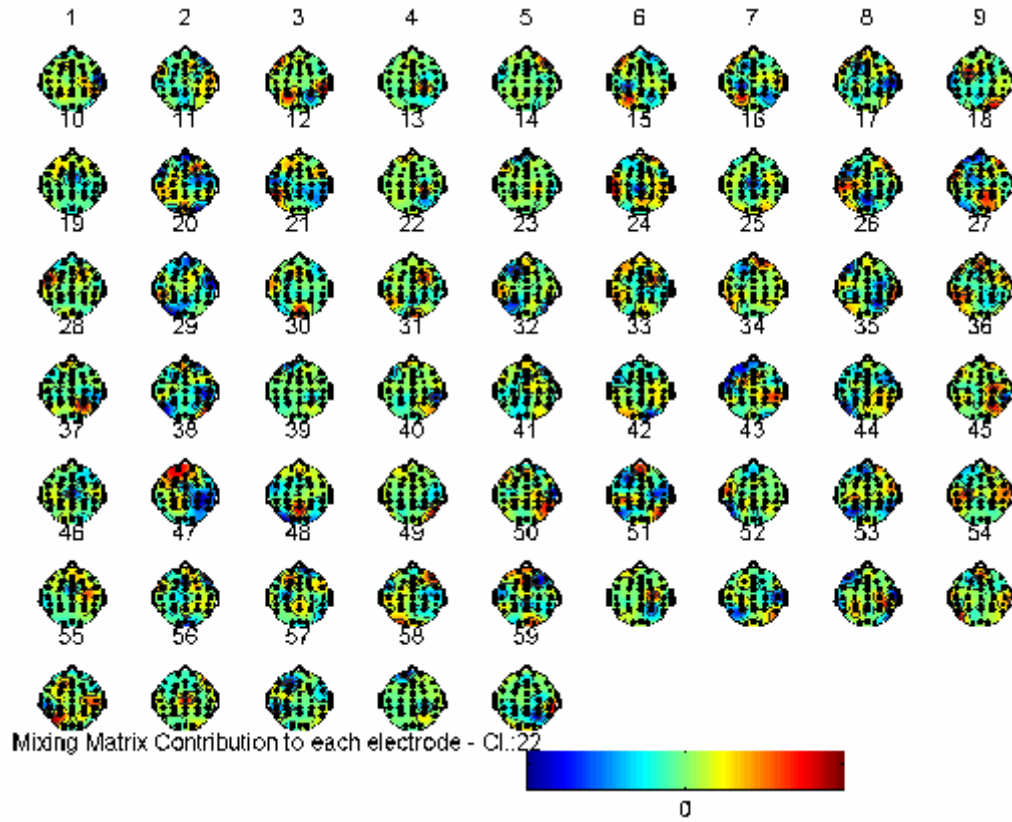


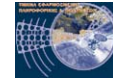




4.3.8.3 Παράδειγμα υπο-ομάδας παρασιτικών πηγών

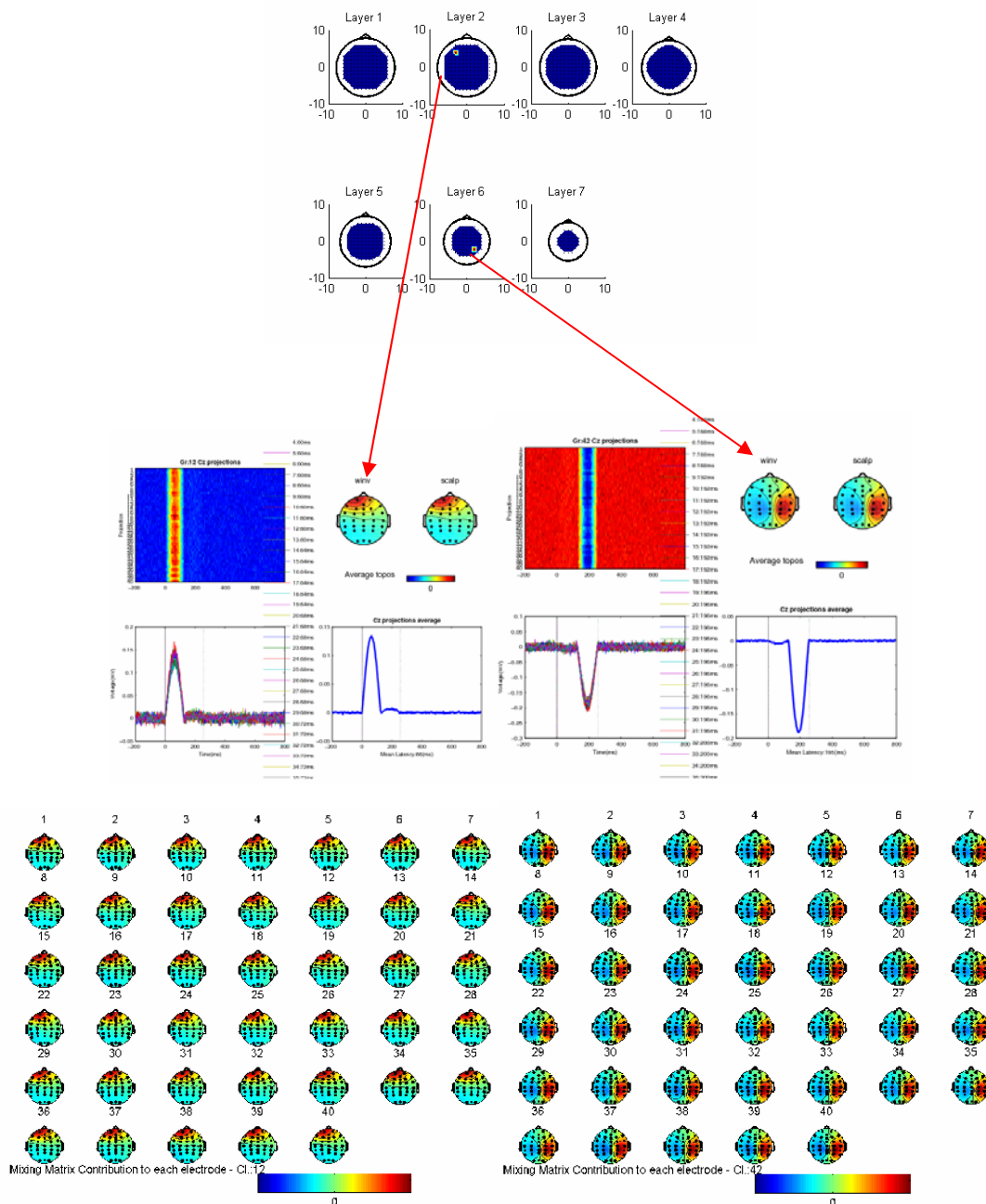


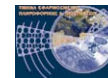




4.4 Σχολιασμός αποτελεσμάτων

Συνολικά η μέθοδος που εφαρμόστηκε με τη χρήση της ICA έδειξε να επιτυγχάνει τον αρχικό στόχο δημιουργώντας ομοιόμορφες ομάδες από συνιστώσες που περιέχουν την αρχική πληροφορία. Παρακάτω, όπως φαίνεται στο σχεδιάγραμμα παρουσιάζεται ένα δείγμα από δυο ομάδες με όμοιους κρανιακούς χάρτες και δίπολα που αντιστοιχούν στις αντιστοιχες πηγές βαση των οποίων δημιουργήθηκαν τα τεχνητά σήματα του πειράματος.





5 Κριτική

Οι στόχοι της εργασίας επιτεύχθηκαν, χωρίς να παρουσιαστούν ιδιαίτερα προβλήματα. Όπως αναφέρθηκε στο κεφ.3.5 βασικά κομμάτια της ερέυνας που αφορούν την αποσυνέλιξη των σημάτων είναι υλοποιημένα από το Salk Institute.

Οι αλγόριθμοι και ο σχετικός κώδικας της μεθόδου εξόρυξης EPs αναπτύχθηκε στο πρόγραμμα MatLab. Η χρήση της σκριπτογλώσσας του MatLab δεν αποτέλεσε πρόβλημα, διότι υπήρχε ήδη η απαραίτητη σχετική εμπειρία και η υλοποίηση είχε διάρκεια ενός μηνός.

Για την απόκτηση της απαραίτητης τεχνογνωσίας συνέβαλε η επιστημονική επίβλεψη του Μ.Βούρκα, με τον οποίο συνεργάστημε αμοιβαία. Για την απόκτηση ενός υψηλού επιπέδου γνώσης για την υλοποίηση της εργασίας χρειάζονται περίπου 3 μήνες.

Μια από τις πιο βασικές προϋποθέσεις για την επίτευξη αυτής της εργασίας είναι το ερευνητικό πνεύμα πέρα από τη σχετική τεχνογνωσία.

Ένα βασικό από τα προβλήματα που αντιμετωπίστηκαν κατά τη διάρκεια της εργασίας αποτέλεσε η έλλειψη ικανοποιητικών δεδομένων. Έτσι λοιπόν, η ανάλυση εφαρμόστηκε σε τεχνητά δεδομένα.

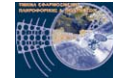


6 Επίλογος

Ολοκληρώνοντας, σύμφωνα με τον αρχικό στόχο επετεύχθει η εξαγωγή των EPs με τη χρήση αποσυνέλιξης σημάτων βάση της τεχνικής ICA, δημιουργώντας ομοιόμορφες ομάδες ως τους κρανιακούς χάρτες των ανεξάρτητων συνιστωσών. Για τα αποτελέσματα χρησιμοποιήθηκε ο αλγόριθμος “tunica” από το Salk Institute, ο οποίος βασίζεται στη θεωρία μεγιστοποίησης της πληροφορίας ενσωματώνοντας την θεωρία της BSS.

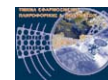
Οστόσο, παρά την επιτυχή υλοποίηση η εν λόγω μεθοδολογία θα μπορούσε να επεκταθεί με την προσθήκη του εντοπισμού των διπολικών πηγών (βλ. Κεφ.8.2) και την ομαδοποίηση των θέσεων των διπόλων, καθώς και ομαδοποίηση των φασματικών χαρακτηριστικών των συνιστωσών.

Η ανάλυση πραγματικών δεδομένων έχει σαν αποδέκτη τους γιατρούς οι οποίοι είναι οι αρμόδιοι να σχολιάσουν τη χρησιμότητα των αποτελεσμάτων της μελέτης και να βγάλουν χρήσιμα συμπεράσματα για τους εξεταζόμενους ασθενείς τους. Η ανάπτυξη και η βελτίωση μίας επιτυχημένης διαδικασίας επεξεργασίας όπως η συγκεκριμένη εργασία μπορεί να εξελιχθεί σε ένα χρήσιμο εργαλείο στα χέρια της νευροεπιστήμης.



7 Βιβλιογραφία

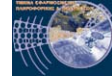
- [1] S.-I. Amari and A. Cichocki. Adaptive blind signal processing—neural network approaches. *Proceedings of the IEEE*, 86(10):2026–2048, 1998.
- [2] F. R. Bach and M. I. Jordan. Kernel independent component analysis. *Journal of Machine Learning Research*, 3:1–48, 2002.
- [3] A. J. Bell and T. J. Sejnowski. A non-linear information maximization algorithm that performs blind separation. In *Advances in Neural Information Processing Systems 7*, pages 467–474. The MIT Press, Cambridge, MA, 1995.
- [4] A.J. Bell and T.J. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7:1129–1159, 1995.
- [5] A. Belouchrani and M. Amin. Blind source separation based on time-frequency signal representations. *IEEE Trans. on Signal Processing*, 46(11):2888–2897, 1998.
- [6] A. Belouchrani, K. Abed Meraim, J.-F. Cardoso, and E. Moulines. A blind source separation technique based on second order statistics. *IEEE Trans. on Signal Processing*, 45(2):434–444, 1997.
- [7] R. Boscolo, H. P. Pan, and V. P. Roychowdhury. Independent component analysis based on nonparametric density estimation. *IEEE Transactions on Neural Networks*, 15:55–65, 2004.
- [8] J.-F. Cardoso. On the performance of orthogonal source separation algorithms. In *Proc. EUSIPCO*, pages 776–779, Edinburgh, Ireland, 1994.
- [9] J.-F. Cardoso. Jacobi angles for simultaneous diagonalization. *SIAM J. Mat. Anal. Appl.*, 17:161–164, 1996.
- [10] J.-F. Cardoso. Infomax and maximum likelihood for source separation. *IEEE Letters on Signal Processing*, 4:112–114, 1997.
- [11] J.-F. Cardoso. High-order contrasts for independent component analysis. *Neural Computation*, 11(1):157–192, 1999.
- [12] S. Choi, S. Amari, A. Cichocki, and R. Liu. Natural gradient learning with a nonholonomic constraint for blind deconvolution of multiple channels, 1999.
- [13] P. Comon. Contrasts for multichannel blind deconvolution. *Signal Processing Letters*, 3(7):209–211, 1996.



- [14] P. Comon, C. Jutten, and J. Héroult. Blind separation of sources, Part I: an adaptive algorithm based on neuromimetic architecture. *Signal Processing*, 24:1–10, 1991.
- [15] P. Comon, C. Jutten, and J. Héroult. Blind separation of sources, Part II: Problems statement. *Signal Processing*, 24:11–20, 1991.
- [16] A. Delorme and S. Makeig. Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134:9–21, 2004.
- [17] J. H. Friedman and J. W. Tukey. A projection pursuit algorithm for exploratory data analysis. *IEEE Trans. of Computers*, c-23(9):881–890, 1974.
- [18] S. Harmeling, A. Ziehe, M. Kawanabe, B. Blankertz, and K. Müller. Nonlinear blind source separation using kernel feature spaces, 2001.
- [19] A. Holobar, C. Févotte, C. Doncarli, and D. Zazula. Single autoterms selection for blind source separation in time-frequency plane. In *Proceeding of 11th EUSIPCO, Toulouse, France*, 2002.
- [20] A. Hyvärinen. Independent component analysis for time-dependent stochastic processes. In *Proc. Int. Conf. on Artificial Neural Networks (ICANN'98)*, pages 135–140, Skövde, Sweden, 1998.
- [21] A. Hyvärinen. Independent component analysis in the presence of gaussian noise by maximizing joint likelihood. *Neurocomputing*, 22:49–67, 1998.
- [22] A. Hyvärinen. New approximations of differential entropy for independent component analysis and projection pursuit. In *Advances in Neural Information Processing Systems*, volume 10, pages 273–279. MIT Press, 1998.
- [23] A. Hyvärinen. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans. on Neural Networks*, 10(3):626–634, 1999.
- [24] A. Hyvärinen. Fast independent component analysis with noisy data using gaussian moments. In *Proc. Int. Symp. on Circuits and Systems*, pages V57–V61, Orlando, Florida, 1999.
- [25] A. Hyvärinen. Complexity pursuit: Separating interesting components from time-series. *Neural Computation*, 13, 2001.
- [26] E. G. Learned-Miller and J. W. Fisher III. Independent component analysis using spacings estimates of entropy. *Journal of Machine Learning Research*, 4:1271–1295, 2003.



- [27] T.-W. Lee, M. Girolami, and T. J. Sejnowski. Independent component analysis using an extended infomax algorithm for mixed sub-gaussian and super-gaussian sources. *Neural Computation*, 11(2):417–441, 1999.
- [28] Klaus-Robert Müller, Petra Philips, and Andreas Ziehe. Jade td : Combining higher-order statistics and temporal information for blind source separation (with noise).
- [29] L. Molgedey and H. G. Schuster. Separation of a mixture of independent signals using time delayed correlations. *Physical Review Letters*, 72:3634–3636, 1994.
- [30] K.-R. Müller, P. Philips, and A. Ziehe.: Combining higher-order statistics and temporal information for blind source separation (with noise). In *Proc. Int. Workshop on Independent Component Analysis and Signal Separation (ICA'99)*, pages 87–92, Aussois, France, 1999.
- [31] M. Ohata and K. Matsuoka. Stability analysis of information-theoretic blind separation algorithms in the case where the sources are nonlinear processes. *IEEE Transactions on Signal Processing*, 50(1):69–77, 2002.
- [32] P. Pajunen. *Extensions of Linear Independent Component Analysis: Neural and Information-theoretic Methods*. PhD thesis, Helsinki University of Technology, 1998.
- [33] P. Pajunen and J. Karhunen. A maximum likelihood approach to nonlinear blind source separation. In *Proceedings of the 1997 Int. Conf. on Artificial Neural Networks (ICANN'97)*, pages 541–546, Lausanne, Switzerland, 1997.
- [34] D.-T. Pham and P. Garat. Blind separation of mixture of independent sources through a quasi-maximum likelihood approach. *IEEE Trans. on Signal Processing*, 45(7):1712–1725, 1997.
- [35] A. Samarov and A. Tsybakov. Nonparametric independent component analysis. *Bernoulli*, 10:565–582, 2004.
- [36] J. Särelä and H. Valpola. Denoising source separation. Technical report, Neural Networks Research Centre, Helsinki University of Technology, 2004.
- [37] J. Särelä and R. Vigario. Overlearning in marginal distribution-based ica: Analysis and solutions. *Machine Learning Research*, 4:1447–1469, 2003.
- [38] H. Stögbauer, A. Kraskov, S. A. Astakhov, and P. Grassberger. Least dependent component analysis based on mutual information. *Bioinformatics*, 2003. Submitted.
- [39] O. Vasicek. A test for normality based on sample entropy. *Journal of the Royal Statistical Society, Series B*, 38:54–59, 1976.



- [40] A. Ziehe and K.-R. Müller. TDSEP—an efficient algorithm for blind separation using time structure. In *Proc. Int. Conf. on Artificial Neural Networks (ICANN'98)*, pages 675–680, Skövde, Sweden, 1998.
- [41] A. Ziehe, G. Nolte, and K.-R. Müller. A Fast algorithm for joint diagonalization with non-orthogonal transformation and its application to blind source separation. *Machine Learning Research*, 5:777:800, 2004.
- [42] Arnaud Delorme, Scott Makeig. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent components analysis. In *Journal of Neuroscience Methods* 134 (2004), 9-21.
- [43] Tzyy-Ping Jung, Scott Makeig, Martin J, McKeown, Anthony J. Bell, Te-Won Lee, and Terrence J. Sejnowski. Imaging Brain Dynamics Using Independent Component Analysis. In *Proceedings of the IEEE, VOL. 89, No. 7, July 2001*.
- [44] A. Hyvärinen and E. Oja. Independent Component Analysis: Algorithms and Applications. *Neural Networks*, 13(4-5):411:430, 2000
- [45] Lucas C. Parra, Clay D. Spence, Adam D. Gerson, and Paul Sajda. Recipes for the linear analysis of EEG. *Elsevier, NeuroImage* 28, 326-341, 2005.
- [46] Nitish V. Thakor and Shanbao Tong. Advances in Quantitative Electroencephalogram Analysis Methods. *Annu. Rev. Biomed. Eng.* 6:453-495, 2004.
- [47] Nick Yeung, Rafal Bogacz, Clay B. Holroyd, and Jonathan D. Cohen. Detection of synchronized oscillators in the electroencephalogram: An evaluation of methods. *Blackwell Publishing Inc. Psychophysiology*, 41, 822-832, 2004.
- [48] Ernesto Pereda, Rodrigo Quiñan Quiroga, Joydeep Bhattacharya. Nonlinear multivariate analysis of neurophysiological signals. *Elsevier, Progress in Neurobiology*, 77, 1-37, 2005.



8 Παρατήματα

8.1 Δείγμα κώδικα υλοποίησης

Ο κώδικας, εκτός των βιβλιοθηκών του MatLab, που έχει χρησιμοποιηθεί και αναπτυχθεί συνολικά για την επίτευξη των αποτελεσμάτων είναι πάνω από 5MB και είναι αδύνατο να συμπεριληφθεί στο κεφάλαιο, για αυτό το λόγω παρακάτω παρατείνεται ένα δείγμα εκτέλεσης του:

```

if 1
parameters;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

for i=1:length(subjects)
    convertUserDataToToolboxFormat (subjects(i), parts(i), TrialTypeIdentifier,
    directoryOfRawData, ...
    directoryOfFormattedData, ICAMethod, whitenData)
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
EP = applyICA2 (subjects, parts, TrialTypeIdentifier, directoryOfFormattedData,
directoryOfICAResults);

end

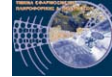
%% _____ Activations, Mixing Matrices, Cz Projections _____ %%

data = EP.X;
[chans frames epochs] = size(data);
chanlist = projectionElectrode;
t = [t1 t2]/1000;
xmin = - t(1);
xmax = ((frames-1)/srate) + xmin ;

for i = 1 : epochs
    weights(:,:,i) = cell2mat(EP.weights(i));
    sphere(:,:,i) = cell2mat(EP.sphere(i));
    white(:,:,i) = cell2mat(EP.whitened(i));
    wm(:,:,i) = cell2mat(EP.WM(i));
    w(:,:,i) = cell2mat(EP.W(i)); % unmixing
matrix * whitening matrix
    activations(:,:,i) = icaact(data(:,:,i), w(:,:,i), mean(data(:,:,i))');%
activations
    winv(:,:,i) = inv( w(:,:,i) ); % mixing
matrix
    for j = 1:chans
        chanproj(j+chans*(i-1),:) = winv(chanlist,j,i)*activations(j,:,i); %
projections on Cz
    end
end
if 1
    eloc = readlocs( 'simLoc31.locs', 'filetype', 'loc');
    pre_stimulus = (t(1)*frames)/(frames/srate);
    post_stimulus = (t(2)*frames)/(frames/srate)+1;
    winlen = 3 * frames/srate;
    spacing = 20;

    if 1
        [dataout rmbase_mean] =
rmbase(activations,frames,[1:pre_stimulus,post_stimulus:frames]);

```



```

    dataout = reshape(dataout,chans,frames,epochs);
    eegplot(activations, 'srate', srate, 'title','Removed reference channel &
baseline', 'dispchans',chans, 'winlength',winlen, 'submean', 'off', 'eloc_file',eloc, 'spaci
ng', spacing, 'data2', dataout, 'limits', [xmin*1000 xmax*1000])
    title('Remove Baseline', 'FontSize', 6)
    print( '-noui', '-dpasc2', '-append', report);
    activations = dataout;
end
end

for i = 1:epochs
    act([1:chans]+chans*(i-1),:) = activations(:, :, i);           % this because of find()
    which return an index vector
    for j = 1:chans
        chanproj(j+chans*(i-1),:) = winv(chanlist,j,i)*activations(j, :, i); %
projections on Cz
    end
end

% Columns of mixing matrices
winv2 = reshape(winv,chans,chans*epochs);
winv2 = winv2';

selections = zeros(chans*epochs,1);

%% _____ Maximum Variance between [t1 t2] _____ %%

% maxframe: frame from ALL component projected on all electrodes
nics = chans;
if 1
    for i = 1 : epochs
        [windex(:,i),maxvar(:,i),maxframe(:,i),maxepoch(:,i),maxmap(:, :, i)] =
EEGcompsoortV2(data(:, :, i), w(:, :, i), 1, frames, nics, [1:chans]);
    end
end

% Reorder again maxframe index to the first order
for i = 1 : epochs
    maxframe(windex(:,i),i) = maxframe(:,i);
end

index = find( t(1)*srate < maxframe(:) & maxframe(:) < t(2)*srate );

selections(index,1) = 1;

clear maxmap maxvar windex maxepoch msg msg1 nics

%% _____ Zero Crossing _____ %%

index = find(selections(:,1));

% Compute zero-crossing mean+std for all activations
% -----
disp('Count number of times signal crosses zero in a single time step')
[zc, zc_mean, zc_std ] = zero_crossing(act(index,pre_stimulus:post_stimulus,:));
%show2D(zc,'Cz projections','Epochs','Zero Crossings')
%print( '-noui', '-dpasc2', '-append', report);
%zc = reshape(zc,chans*epochs,1);

% Visualize zero crossing rejected samples
q=find(zc >= zc_mean + zc_std);
q=index(q);
for i = 1:40:length(q)-40
    figure,plotdata(act(q(i:i+39),:),frames,[[xmin xmax]*1000
datlim(act(q(i:i+39),:))*0.4], 'Zero crossing rejected samples')
    print( '-noui', '-dpasc2', '-append', report);
end
q=find(zc < zc_mean + zc_std);
q=index(q);
selections(q,2)=1;

clear zc zc_mean zc_std q weights_size tmp

%% _____ Extract Clustering Features _____ %%

```



```

index = find(selections(:,2));
framestep = (1000/srate);
for i = 1:length(index)
    peaks_selected(i,:) = [ maxframe(index(i))*framestep
chanproj(index(i),maxframe(index(i)))];
end

figure('Visible','off'),plot(peaks_selected(:,1), abs(peaks_selected(:,2)),'.b'),grid
on
xlabel('Latency(ms)'),ylabel('Amplitude(mV)'),title(strcat('Extracted peaks within
:',num2str(abs(xmin)),'-',num2str(t(2)),' sec'),'FontWeight','bold')
print( '-noui', '-dpsc2', '-append', report);

features = [peaks_selected(:,1) abs(peaks_selected(:,2))];
%features = normaliz_byMax(features);
%features = normaliz_byNorm2(features);

%% _____ 1st Clustering _____ %%

% Cluster peaks & latencies & energy
% -----
disp('Primary Clustering: Unsupervised k-means Clustering for Cz projections peaks')
clust1_num = 5;
[IDX,C,sumd,D,outliers] = robust_kmeans(features,clust1_num,5,10);
%[clust1_num, IDX]= doUnsupervisedKmeansClustering (features, clust1_num, 5);
figure('Visible','off'),
[s,h] = silhouette(features,IDX,'sqeuclid');
str{1}='Primary Clustering: Unsupervised k-means';
str{2}=strcat('Clustered Cz projections
peaks:',num2str(length(peaks_selected(find(IDX)))));
str{3}=strcat('Outlier peaks from k-means:',num2str(length(find(IDX==0))));
title(str,'FontWeight','bold')
for i=1:clust1_num
    str2{i} = strcat('Cl.',num2str(i),' #',num2str(length(find(IDX==i))) );
end
text(1.15,.5,str2,'FontSize',7)
print( '-noui', '-dpsc2', '-append', report);

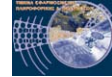
selections(index,3) = IDX;
tmp = find( IDX );
peaks_selected = peaks_selected(tmp,:);
IDX = selections(find(selections(:,3)),3);

if 1
% Group statistics for centroid distances
[GRP1.mean GRP1.sem GRP1.numel GRP1.gname GRP1.std GRP1.var GRP1.meanci GRP1.predci] =
grpstats(min(D'), IDX, {'mean', 'sem', 'numel', 'gname', 'std', 'var', 'meanci',
'predci'},0.05);
[GRP1.mean GRP1.sem GRP1.numel GRP1.gname GRP1.std GRP1.var GRP1.meanci GRP1.predci] =
grpstats(min(D'), IDX, {'mean', 'sem', 'numel', 'gname', 'std', 'var', 'meanci',
'predci'},0.05);
figure('Visible','off'),
subplot(2,1,1),bar(GRP1.numel),xlabel('Clusters'),ylabel('Number of
components'),title('Number of Components/Cluster','FontWeight','bold')
subplot(2,1,2),bar([GRP1.mean GRP1.sem
GRP1.std],1),colormap('summer'),legend('Mean','Standard error of the mean','Standard
deviation','Location',[.82 .42 .05 .1])
legend('boxoff'),xlabel('Clusters'),ylabel('Statistic Values'),title('Centroid
Distances/Cluster','FontWeight','bold')
hold on,errorbar(GRP1.mean,GRP1.std,'b'),hold on,errorbar(GRP1.mean,GRP1.sem,'r')
print( '-noui', '-dpsc2', '-append', report);
end
% Visualize Clustered Peaks
figure('Visible','off'),
scatter(peaks_selected(:,1),abs(peaks_selected(:,2)),15,IDX,'filled')
ylabel('Amplitude(mV)'),xlabel('Latency(ms)'),title(strcat('Peaks within
:',num2str(abs(xmin)),'-',num2str(t(2)),' sec - removed outliers from k-
means'),'FontWeight','bold')
print( '-noui', '-dpsc2', '-append', report);

save(strcat(datapath,savemat));

close all

```

```

%% _____ 2nd Clustering _____ %%
clear str
clust2_num = 5;
for i = 1:clust1_num
    index = find( selections(:,3) == i);

    [IDX,C] = neural_net(winv2(index,:),clust2_num);
    %[IDX,C,sumd,D,outliers] = robust_kmeans(winv2(index,:),clust2_num,3,10);
    %[IDX,C,sumd,D] =
kmeans(winv2(index(find(cols))),clust2_num,'replicates',30,'emptyaction','drop');

    figure('Visible','off')
    [s,h] = silhouette(winv2(index,:),IDX,'sqeuclid');
    str{1}=strcat('Secondary Clustering: Neural Net Clustering - Subcluster:
#',num2str(i));
    %str{1}=strcat('Secondary Clustering: Unsupervised k-means - Subcluster:
#',num2str(i));
    str{2}=strcat('Clustered mixing matrices columns:',num2str(length( find(IDX) )));
    %str{3}=strcat('Outlier columns from k-means:',num2str(length(index) - length(
find(IDX) )));
    title(str,'FontWeight','bold')
    for j=1:clust2_num
        str2{j} = strcat('Cl.',num2str(j),' #',num2str(length(find(IDX==j))) );
    end
    text(1.15,.5,str2,'FontSize',7)
    print( '-noui', '-dpasc2', '-append', report);
    selections(index,4) = IDX+10*i;
    %dist(index(find(cols)),1) = min(D)';

    if 0
    % Group statistics for centroid statistics
    [GRP2(i).mean GRP2(i).sem GRP2(i).numel GRP2(i).gname GRP2(i).std GRP2(i).var
GRP2(i).meanci GRP2(i).predci] = grpstats(min(D'),selections(index(find(IDX)),4),
{'mean', 'sem', 'numel', 'gname', 'std', 'var', 'meanci', 'predci'},0.05);
    figure('Visible','off')
    subplot(2,1,1),bar(GRP2(i).numel),xlabel('Clusters'),ylabel('Number of
components'),title(strcat('Number of
Components/Cluster.',num2str(i),'FontWeight','bold')
    subplot(2,1,2),bar([GRP2(i).mean GRP2(i).sem
GRP2(i).std],1),colormap('summer'),legend('Mean','Standard error of the
mean','Standard deviation','Location',[.82 .42 .05 .1])
    legend('boxoff'),xlabel('Clusters'),ylabel('Statistic Values'),title('Centroid
Distances/Cluster','FontWeight','bold')
    hold on,errorbar(GRP2(i).mean,GRP2(i).std,'b'),hold
on,errorbar(GRP2(i).mean,GRP2(i).sem,'r')
    print( '-noui', '-dpasc2', '-append', report);
    end
end
close all

%% _____ Clusters Manipulation _____ %%

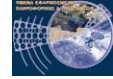
% Separate +/- peaks
% -----
disp('Separating +/- peaks...')
index = find( selections(:,3) ); % because of
"peaks_selected" matrix siz
negative = find( peaks_selected(:,1) < 0 ); % index for
negative peaks % check them as
selections(index(negative),4) = - selections(index(negative),4);
negative groups

[G,GN]=grp2idx(selections(find(selections(:,4)),4));
for i = 1 : length(GN)
    groups(i,1)=str2num(GN{i}); % final groups indexing
end

%% Compute the scalp maps at each maximum variance latency

close all
scalp_Cz_maxvar = zeros(size(winv2));
ind = find( selections(:,3) );
index = find(selections(ind,4));
for i = 1:length(index)

```



```

        latency = peaks_selected(index(i),1)/(1000/srate);           % ms latency to frames
        scalp_Cz_maxvar(ind(index(i)), :) =
winv2(ind(index(i)), :) * act(ind(index(i)), latency);
end

%% _____ Plot Topographies to choose meaningful Clusters _____ %%

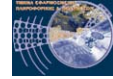
close all
comps = [];
grpidx = [];

dipoles= 0;

for i = 1 : length(groups)
    plotcomps = find(selections(:,4) == groups(i));           % plot comps sort by
groups
% -----
    if dipoles == 1
        TMP.posxyz = [0 0 0];
        TMP.momxyz = [0 0 0];
        TMP.rv      = [0];
    end
    if length(plotcomps)>1
        % ERP average
        erp_ave = sum(chanproj(plotcomps, :))/length(plotcomps);
        % AveWinv
        mean_winv = [sum(winv2(plotcomps, :))/length(plotcomps)]';
        % AveScalp
        mean_scalp = [sum(scalp_Cz_maxvar(plotcomps, :))/length(plotcomps)]';
        % AveSource
        if dipoles == 1
            for j = 1:length(plotcomps)
                TMP.posxyz = SOURCES(plotcomps(j)).posxyz + TMP.posxyz;
                TMP.momxyz = SOURCES(plotcomps(j)).momxyz + TMP.momxyz;
                TMP.rv      = SOURCES(plotcomps(j)).rv + TMP.rv;
            end
            TMP.posxyz = TMP.posxyz /length(plotcomps);
            TMP.momxyz = TMP.momxyz /length(plotcomps);
            TMP.rv      = TMP.rv      /length(plotcomps);
        end
        % MeanLat
        ind = find(selections(:,3));
        ind = find(selections(ind,4) == groups(i));
        mean_lat = round(sum(peaks_selected(ind,1))/length(plotcomps)+xmin*1000);
        [s index] = sort(peaks_selected(ind,1));
    elseif length(plotcomps)==1
        ind = find(selections(:,3));
        ind = find(selections(ind,4) == groups(i));
        if dipoles == 1
            TMP = SOURCES(plotcomps);
        end
        erp_ave=chanproj(plotcomps, :);
        mean_lat=peaks_selected(ind,1)+xmin*1000;
        mean_winv=winv2(plotcomps, :)';
        mean_scalp=scalp_Cz_maxvar(plotcomps, :)';
        [s index] = sort(peaks_selected(ind,1));
    end
    legends = {};
    for j = 1 :length(plotcomps)
        legends{j} =
strcat(num2str(j), ':', num2str(peaks_selected(ind(index(j)),1)+xmin*1000), 'ms') ;
    end
    str = strcat('Mean Latency:', num2str(mean_lat), ', Cl.: ', num2str(groups(i)));
    figure, time_vector = xmin*1000:(1000*(frames/srate))/frames:xmax*1000;

h1=subplot(2,2,1); imagesc(time_vector, [], chanproj(plotcomps(index), :)); t1=title(strc
at('Gr:', num2str(groups(i)), ' Cz projections')); set(gca, 'YTick', [1 :
length(plotcomps)]), y1=ylabel('Projection'));
    if dipoles == 1
        str1 = strcat('POSxyz: ', num2str([floor(TMP.posxyz)]), ' /
RV:', num2str([floor(TMP.rv*100)]), '%');
        h4=subplot(2,2,2); dipmap([mean_winv, mean_scalp], eloc, 0, str1, 0, ['winv
'; 'scalp'], 1, [TMP; TMP])
    else

```



```

        h4=subplot(2,2,2); compmap([mean_winv,mean_scalp],eloc,0,'Average
topos',0,['winv ':'scalp'],1);
    end
    h3=subplot(2,2,4); plot(time_vector,erp_ave,'LineWidth',2),Y =
get(gca,'YLim'); line(0,Y(1):diff(Y)/400:Y(2)), line([t(2)-
t(1)]*1000,Y(1):diff(Y)/80:Y(2)),x1=xlabel(strcat('Mean
Latency:',num2str(mean_lat),'(ms)')); t2=title('Cz projections average');
%set(gca,'YLim',datlim(chanproj(plotcomps,:)));
    h2=subplot(2,2,3); plot(time_vector,chanproj(plotcomps(index),:)); h =
legend(legends,'Location',[.45 .42 .05 .1]); legend('boxoff'),Y =
get(gca,'YLim'); line(0,Y(1):diff(Y)/400:Y(2)), line([t(2)-
t(1)]*1000,Y(1):diff(Y)/80:Y(2)),x2=xlabel('Time(ms)'); y2=ylabel('Voltage(mV)');

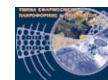
set([h1,h2,h3,h,x1,x2,y1,y2],'FontSize',7); set([t1,t2],'FontWeight','Bold','FontSize',
,8);
    print( '-noui', '-dpSC2', '-append', report);
    % -----
    % Labels of grouped & sorted components: *** plotcomps(index) ***
    comp = [];
    for jj = 1:size(plotcomps)
        c = mod(plotcomps(index(jj)),chans); if c==0, c=chans; end
        comp(jj).labels = strcat(' Tr:',num2str(ceil(plotcomps(index(jj))/chans)
),' Co:',num2str(c));
        comps(size(comps,2)+1).labels =
strcat('Gr:',num2str(groups(i)),comp(jj).labels);
    end
    grpidx = [ grpidx; plotcomps(index) ]; % Total index of Grouped & Sorted
Components
    % -----
    tmpqqq = []; tmpqqq(:,1) = chanproj(plotcomps(index),:); tmpqqq(:,2) =
chanproj(plotcomps(index),:);
    eegplot(tmpqqq, 'srate',
srate,'title','Projections','dispchans',length(plotcomps),'winlength',frames/srate,'su
bmean','off','spacing',max(abs(datlim(tmpqqq))),'eloc_file',comp,'limits',[xmin*1000
xmax*1000]),title(strcat('Gr:',num2str(groups(i)),' projections'),'FontSize',6)
%figure,plotdata(tmpqqq,frames,[xmin xmax]*1000
datlim(tmpqqq),strcat('Gr:',num2str(groups(i)),' Cz projections'),comp)
    print( '-noui', '-dpSC2', '-append', report);
    % -----
    if dipoles == 1
        dipplot(
SOURCES(plotcomps(index)),'coordformat','spherical','normlen','on','mesh','off','num',
'on','projlines','on');
        view([45 30])
        print( '-noui', '-dpSC2', '-append', report);
    end
    % -----
    str = strcat('Mixing Matrix Contribution to each electrode - Cl.:
',num2str(groups(i)));
    if dipoles == 1

figure,dipmap(win2(plotcomps(index),:),'eloc,0,str,0,0,1,SOURCES(plotcomps(index)))
    else
        figure,compmap(win2(plotcomps(index),:),'eloc,0,str,0,0,1)
    end
    print( '-noui', '-dpSC2', '-append', report);
    % -----
    str = strcat('Scalp maps at MaxVar from Cz projections - Cl.:
',num2str(groups(i)));
    if dipoles == 1

figure,dipmap(scalp_Cz_maxvar(plotcomps(index),:),'eloc,0,str,0,0,1,SOURCES(plotcomps(
index)))
    else
        figure,compmap(scalp_Cz_maxvar(plotcomps(index),:),'eloc,0,str,0,0,1)
    end
    print( '-noui', '-dpSC2', '-append', report);
    % -----
    close all
end

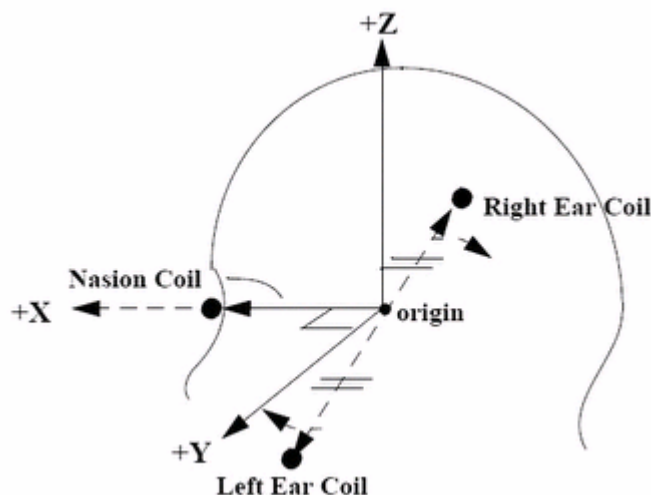
clear Y acs alpha ans comp filtorder h h1 h2 h3 h4 i jj j ind index
clear latency plotcomps legends dipoles ind index mean_lat report
clear str x1 x2 y1 y2 tmp tmp_b_w tmp_retain tmpindex
%save(strcat(datapath,savemat));

```



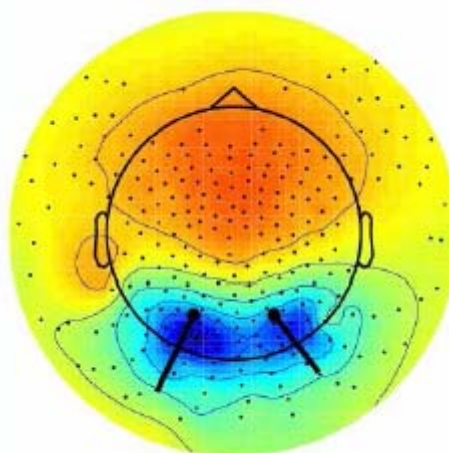
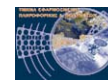
8.2 Προσέγγιση διπολικών πηγών

Ένα σημαντικό εμπόδιο στην οπτικοποίηση μακροσκοπικών εγκεφαλικών δυναμικών, για EEG δεδομένα, είναι η απροσδιόριστη φύση του προβλήματος αντιστροφής.



Σχ. Μια αναπαράσταση του εγκεφαλικού συστήματος συντεταγμένων και η σχέση του με το σημείο ορόσημο.

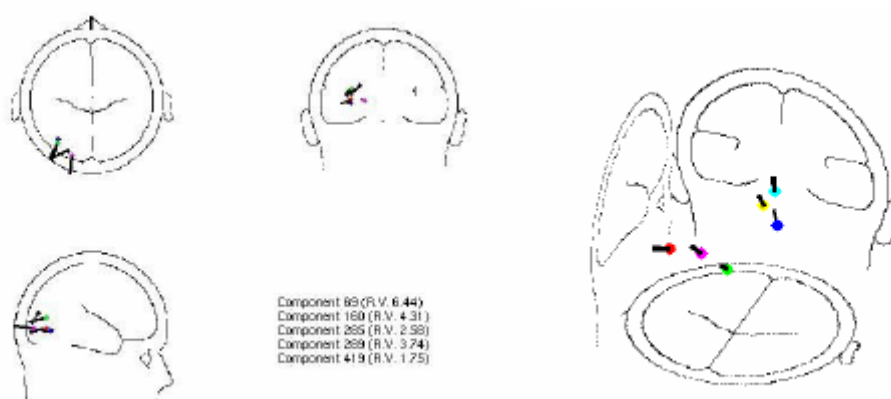
Δηλαδή, μία γνωστή κρανιακή κατανομή για καταγεγραμμένη δραστηριότητα από γνωστά ηλεκτροδία, μπορεί να έχει προκληθεί από οποιάδήποτε εγκεφαλική διπολική κατανομή. Αυτό συμβαίνει γιατί είναι δυνατοί πάρα πολλοί συνδυασμοί ζευγών μεταξύ των εγκεφαλικών περιοχών. Μία υπολογιστική προσέγγιση σαν στόχο έχει να εντοπίσει τα ισοδύναμα δίπολα (κάτι σαν εξαιρετικά μικρές μπαταρίες) των οποίων το άθροισμα των προβολών τους στην επιφάνεια του κρανίου είναι μοιάζει με την καταγεγραμμένη κρανιακή κατανομή (κρανικός χάρτης).



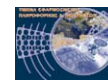
Σχ. Ο εγκεφαλικός χάρτης και τα αντίστοιχα δίπολα σε δισδιάστατη αναπαράσταση.

Έχει αποδειχτεί ότι, πολλές ανεξάρτητες συνιστώσες έχουν κρανικούς χάρτες που σχεδόν τέλεια ταιριάζουν με την προβολή ενός μοναδιαίου ισοδύναμου διπόλου. Αυτή η διατύπωση είναι σύμφωνη με την αναμενόμενη δημιουργία των διπόλων μέσω μερικού συγχρονισμού των διαδικασιών των τοπικών δυναμικών (LPF).

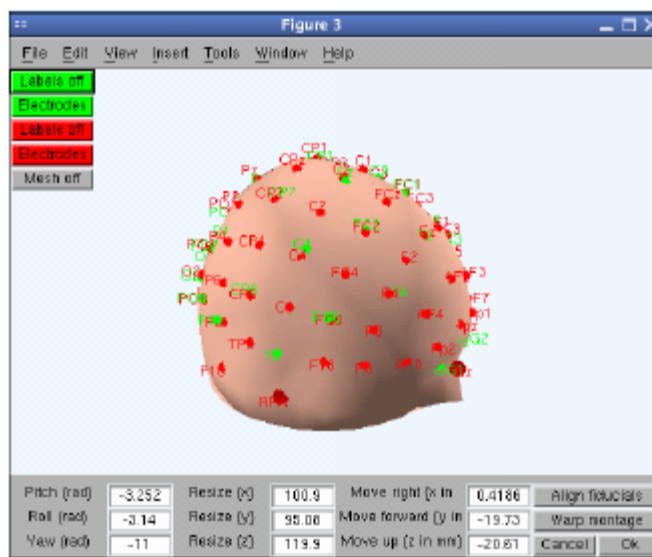
Οι κρανιακοί χάρτες πολλών ICA συνιστωσών είναι συμβατές με τη δημιουργία τους σε ξεχωριστές περιοχές του εγκεφαλικού φλοιού από μερικό συγχρονισμού. Επειδή οι τοπικές συνδέσεις στο φλοιό παρουσιάζουν μεγαλύτερη πυκνότητα από τις μακρύτερου εύρους συνδέσεις, είναι λογικό να υποθέτουμε ότι οι συγχρονισμένες συζεύξεις από νευρωνική δραστηριότητα, που απομονώνονται από την ICA, συνήθως συμβαίνουν εντός μίας απλής εγκεφαλικής περιοχής.



Σχ. Μια αναπαράσταση εντοπισμένων διπόλων .



Για να πραγματοποιηθεί ο εντοπισμός των αυθεντικών πηγών χρησιμοποιείται μία τεχνική μη-γραμμικής βελτιστοποίησης (Scherg 1990) ενός σφαιρικού μοντέλου 4-διαστάσεων (Kavanagh 1978) για την προσαρμογή στο ισοδύναμο διπολικό μοντέλο.



Σχ. Προσαρμογή του Location File(αρχείο συντεταγμένων) στο σφαιρικό εγκεφαλικό μοντέλο.

Εντοπισμός πηγών: Οι θέσεις των ηλεκτροδίων προβάλλονται στην επιφάνεια του κρανίου βάση του σφαιρικού μοντέλου. Ένα σκανάρισμα του πλέγματος εσωτερικά του κρανίου εφαρμόζεται για να υπολογιστεί ένα πρωιμο μοντέλο (δίπολο προς ηλεκτρόδιο) για κάθε θέση του πλέγματος και το συγκρίνει με όλους τους κρανιακούς χάρτες από τις ICA συνιστώσες. Τα δίπολα που διατηρούνται έχουν υπολειμματική διακύμανση (έναντι του αντιστοίχου κρανιακού χάρτη) κάτω από κάποιο όριο (π.χ. 20%).

Σημειώνεται ότι, η πολικότητα των συνιστωσών δεν είναι συγκεκριμένη. Το προϊόν των τιμών του κρανιακού χάρτη από το χάρτη ενεργοποίησης (των προβαλλόμενου συνεισφορών από τις συνιστώσες, δηλ. τις στήλες του πίνακα μείξης) είναι αδιευκρινιστές.