

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ: ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ ΤΗΣ ΠΛΗΡΦΟΡΙΑΣ
ΕΡΓΑΣΤΗΡΙΟ

Διπλωματική Εργασία

της φοιτήτριας του Τμήματος Ηλεκτρολόγων Μηχανικών και
Τεχνολογίας Υπολογιστών της Πολυτεχνικής Σχολής του
Πανεπιστημίου Πατρών

ΧΑΝΤΖΙΑΡΑ Ι. ΜΑΡΙΑ

Αριθμός Μητρώου: **6131**

Θέμα

**«Κατασκευή συστήματος ταυτόχρονης
αναγνώρισης ομιλίας»**

Επιβλέπων

ΔΕΡΜΑΤΑΣ ΕΥΑΓΓΕΛΟΣ

Αριθμός Διπλωματικής Εργασίας:

Πάτρα, Οκτώβριος 2012

ΠΙΣΤΟΠΟΙΗΣΗ

Πιστοποιείται ότι η Διπλωματική Εργασία με θέμα

**«Κατασκευή συστήματος ταυτόχρονης
αναγνώρισης ομιλίας»**

Της φοιτήτριας του Τμήματος Ηλεκτρολόγων Μηχανικών και
Τεχνολογίας Υπολογιστών

XANTZIAPA I. MARIA

Αριθμός Μητρώου:**6131**

Παρουσιάστηκε δημόσια και εξετάστηκε στο Τμήμα
Ηλεκτρολόγων Μηχανικών και Τεχνολογίας Υπολογιστών στις
...../...../.....

Ο Επιβλέπων

Επίκουρος Καθηγητής
Δερματάς Ευάγγελος

Ο Διευθυντής του Τομέα

Καθηγητής
Φακωτάκης Νικόλαος

Αριθμός Διπλωματικής Εργασίας:

Θέμα:

«Κατασκευή συστήματος ταυτόχρονης αναγνώρισης ομιλίας»

Φοιτήτρια:

Χαντζιάρη Ι. Μαρία

Επιβλέπων:

Δερματάς Ευάγγελος

Περίληψη

Σκοπός της παρούσας διπλωματικής εργασίας είναι η δημιουργία ενός συστήματος μίξης ηχητικών σημάτων και προσπάθεια διαχωρισμού τους με βάση τις μεθόδους τυφλού διαχωρισμού σημάτων. Έχοντας ως δεδομένα τα αρχικά σήματα των πηγών γίνεται προσπάθεια, αρχικά μέσω της εφαρμογής της μεθόδου Ανάλυσης Ανεξάρτητων Συνιστωσών (ICA) για την περίπτωση της στιγμιαίας μίξης και στη συνέχεια μέσω της χρήσης αλγορίθμων που στηρίζονται στο μοντέλο παράλληλου παράγοντα (PARAFAC) για την περίπτωση της συνελικτικής μίξης, να προσδιοριστούν τα σήματα των πηγών από τα σήματα μίξης. Επιπλέον, τροποποιώντας τις παραμέτρους του συστήματος που μελετάμε σε κάθε περίπτωση, προσπαθούμε να πετύχουμε τη βέλτιστη απόδοση του διαχωρισμού.

He who knows does not speak, he who speaks does not know

Lao Tzu

Ευχαριστίες

Θα ήθελα να ευχαριστήσω θερμά τον επιβλέποντα καθηγητή μου, Δρ. Ευάγγελο Δερματά, Επίκουρο Καθηγητή στο Πανεπιστήμιο Πατρών, για την καθοδήγησή του κατά τη διάρκεια εκπόνησης της εργασίας αυτής, αλλά και για τον άμεσο και ουσιαστικό τρόπο διδασκαλίας των προπτυχιακών μαθημάτων.

Τον Δρ. Αθανάσιο Κούτρα, για την παροχή των απαραίτητων δεδομένων για την εκπόνηση της εργασίας.

Τον αδερφό μου, για την ουσιαστική βοήθεια και την ψυχολογική στήριξη που μου προσέφερε, τόσο κατά τη διάρκεια της εκπόνησης της διπλωματικής εργασίας, όσο και γενικότερα καθ' όλη τη διάρκεια των σπουδών μου.

Τους γονείς μου, την οικογένειά μου, για όλη τη συμπαράσταση και την στήριξή τους.

Τους φίλους και συμφοιτητές μου, για την κατανόησή τους και την ψυχολογική τους συνδρομή στη διάρκεια εκπόνησης της εργασίας, αλλά και στις δύσκολες και στις ωραίες στιγμές που περάσαμε στα χρόνια των σπουδών μας.

ΠΕΡΙΕΧΟΜΕΝΑ

ΠΕΡΙΛΗΨΗ	5	
ΠΕΡΙΕΧΟΜΕΝΑ	11	
ΕΙΣΑΓΩΓΗ	14	
Κεφάλαιο 1		
Το Σύστημα Αναγνώρισης ομιλίας		
1.1	Εισαγωγή	15
1.2	Ποικιλία της Ομιλίας	15
1.2.1	Διαφορές στην ομιλία του ίδιου ομιλητή	16
1.2.2	Διαφορές στην ομιλία διαφορετικών ομιλητών	16
1.3	Ακουστική και ομιλία	18
1.3.1	Ακουστική δωματίου	18
1.3.2	Μικρόφωνα	19
Κεφάλαιο 2		
Ανάλυση Ανεξάρτητων Συνιστωσών		
2.1	Σκοπός της μεθόδου	21
2.2	Ανάλυση Ανεξάρτητων Συνιστωσών	26
2.2.1	Ορισμός της ICA	26
2.2.2	Ασάφειες της ICA	28
2.2.3	Απεικόνιση της ICA	28
2.3	Περί ανεξαρτησίας	31
2.3.1	Ορισμός και βασικές ιδιότητες	31
2.3.2	Οι ασυσχέτιστες μεταβλητές είναι μόνο μερικώς ανεξάρτητες	32
2.3.3	Γιατί οι Γκαουσιανές μεταβλητές είναι απαγορευμένες	32
2.4	Παραδοχές υπολογισμού της ICA	33
2.4.1	«Η μη-Γκαουσιανή είναι ανεξάρτητη»	33
2.4.2	Μεγέθη της μη προσαρμογής σε Γκαουσιανή κατανομή	35
2.4.2.1	Κύρτωση	35
2.4.2.2	Αρνητική Εντροπία	38
2.4.2.3	Προσεγγίσεις της αρνητικής εντροπίας	40
2.4.3	Ελαχιστοποίηση της αμοιβαίας πληροφορίας	41
2.4.3.1	Αμοιβαία Πληροφορία	41
2.4.3.2	Ορίζοντας την ICA με βάση την αμοιβαία πληροφορία	42
2.4.4	Εκτίμηση της Μέγιστης Πιθανοφάνειας	43
2.4.4.1	Η Πιθανότητα	43
2.4.4.2	Η αρχή της μεγιστοποίησης της εντροπίας (The infomax principle)	43
2.4.4.3	Συσχέτιση με την αμοιβαία πληροφορία	44
2.5	Προεπεξεργασία για την ICA	44
2.5.1	Κεντροθέτηση	44
2.5.2	Αποσυσχέτιση	45
2.5.3	Περαιτέρω προεπεξεργασία	47
2.6	Ο αλγόριθμος FastICA	48

2.6.1	Περιγραφή λειτουργίας του FastICA σε μία μονάδα νευρωνικών δικτύων	48
2.6.2	Περιγραφή λειτουργίας του αλγόριθμου FastICA σε πολλές μονάδες νευρωνικών δικτύων	50
2.6.3	Σχέση FastICA και Μέγιστης Πιθανότητας	51
2.6.4	Ιδιότητες του αλγόριθμου FastICA	51
2.7	Εφαρμογές της ICA	52
2.7.1	Διαχωρισμός των παράσιτων (artifacts) στα δεδομένα MEG	52
2.7.2	Εύρεση κρυμμένων παραγόντων σε οικονομικά δεδομένα	56
2.7.3	Μείωση θορύβου στις φυσικές εικόνες (natural images)	59
2.7.4	Τηλεπικοινωνίες	62

Κεφάλαιο 3

Εφαρμογές της μεθόδου ICA σε πρότυπα σήματα και σήματα φωνής

3.1	Εισαγωγή	63
3.2	Εφαρμογές σε πρότυπα σήματα	63
3.3	Εφαρμογή σε σήματα φωνής εξαρτημένα από τις αποστάσεις των ομιλητών χρησιμοποιώντας τον αλγόριθμο fastICA	66
3.4	Εφαρμογή του fastICA σε ένα σήμα φωνής και μια μελωδία.	68
3.5	Εφαρμογή του fastICA για το διαχωρισμό σημάτων δύο ομιλητών σε ένα δωμάτιο	71

Κεφάλαιο 4

Το πρόβλημα του Υπερ-ορισμένου διαχωρισμού σημάτων

4.1	Εισαγωγή	73
4.2	Υπερ-ορισμένος διαχωρισμός σε στιγμιαία μίξη	74
4.2.1	Το μοντέλο της μίξης	74
4.2.2	Υπάρχουσες μέθοδοι διαχωρισμού	74
4.3	Ορισμός του PSNR	76
4.4	Μετρήσεις-Αποτελέσματα	77

Κεφάλαιο 5

Συνελικτική Μίξη με χρήση του Μοντέλου PARAFAC

5.1	Εισαγωγή	80
5.2	Διατύπωση του προβλήματος	82
5.2.1	Δεδομένα του προβλήματος	82
5.2.2	Υπολογισμός του καναλιού	84
5.3	Σύνδεση με το μοντέλο PARAFAC	85
5.3.1	Αναδιατύπωση του προβλήματος	85
5.3.2	Αναγνωρισιμότητα	86
5.4	Μαζική Υλοποίηση (Batch PARAFAC)	87
5.4.1	Απεικόνιση του διανύσματος σε πίνακα	87
5.4.2	Υπολογισμός του διαχωρισμού PARAFAC	88
5.5	Ασάφειες λόγω βάρθρωσης και αντιμετάθεσης	90
5.5.1	Ασάφεια λόγω βάρθρωσης	90
5.5.2	Ασάφεια λόγω αντιμετάθεσης	91

5.5.3	Σύγκριση μεταξύ των λυτών αντιμετάθεσης	93
5.6	Υπο-ορισμένο πρόβλημα	94
5.7	Υλοποίηση με συνεχή επαναπροσδιορισμό	96

Κεφάλαιο 6

Εφαρμογή του μοντέλου PARAFAC-Μετρήσεις

6.1	Εισαγωγή	98
6.2	Διαδικασία υλοποίησης του αλγορίθμου της συνελικτικής μίξης	98
6.3	Αξιολόγηση της απόδοσης	99
6.4	Υπερ-ορισμένος διαχωρισμός σημάτων σε συνελικτική μίξη	100
6.4.1	Πραγματοποίηση μετρήσεων μεταβάλλοντας τις παραμέτρους του δωματίου	103
6.5	Υπο-ορισμένος διαχωρισμός σημάτων σε συνελικτική μίξη	113

ΠΑΡΑΡΤΗΜΑ	115
------------------	------------

ΒΙΒΛΙΟΓΡΑΦΙΑ	131
---------------------	------------

ΕΙΣΑΓΩΓΗ

Αντικείμενο της παρούσας εργασίας είναι η επεξεργασία μεθόδων Ανάλυσης Ανεξάρτητων Κυρίαρχων Συνιστωσών (Independent Component Analysis) με εφαρμογή σε ηχητικά σήματα. Σκοπός αυτής της μεθόδου είναι η ανάκτηση των ανεξάρτητων κυρίαρχων συνιστωσών (δηλαδή τα σήματα των πηγών) από τα σήματα μίξης.

Αρχικά, πραγματοποιήθηκαν εφαρμογές με τη μέθοδο ICA σε δύο μορφές (πρότυπα, φωνής) σημάτων μέσω MATLAB. Η μέθοδος αυτή εκτελέστηκε με τη χρήση του αλγόριθμου fastICA, ο οποίος δίνει τη δυνατότητα διαχωρισμού των πηγών από τα σήματα μίξης.

Γενικά, αυτή η μέθοδος μπορεί να χρησιμοποιηθεί σε διάφορα πεδία όπως: οπτική απεικόνιση νευρώνων, αναγνώριση προσώπου, πρόβλεψη των τιμών του χρηματιστηρίου, κινητές τηλεφωνικές επικοινωνίες, αναγνώριση χρώματος αντικειμένου, ηλεκτροεγκεφαλογράφημα, επεξεργασία φωνής, κ.τ.λ.

Η εργασία αυτή έχει οργανωθεί στα ακόλουθα κεφάλαια ως εξής:

Στο 1^ο κεφάλαιο, παρουσιάζεται το σύστημα αναγνώρισης ομιλίας.

Στο 2^ο κεφάλαιο, περιγράφεται η έννοια και ο ορισμός της μεθόδου ICA από μαθηματική άποψη καθώς και με απλές εφαρμογές. Επίσης περιγράφεται, ο βασικός αλγόριθμος που χρησιμοποιείται για την επίλυση αυτών των προβλημάτων, ο οποίος ονομάζεται fastICA.

Στο 3^ο κεφάλαιο, δίνονται απλά παραδείγματα της μεθόδου ICA σε μίξη φωνής.

Στο 4^ο κεφάλαιο, θα μελετήσουμε την επίδραση του πλήθους των χρησιμοποιούμενων δεκτών στη βελτίωση της απόδοσης των συστημάτων αναγνώρισης ομιλίας, παρουσιάζοντας το πρόβλημα του υπερ-ορισμένου τυφλού διαχωρισμού σημάτων (Overdetermined Blind source Separation).

Στο 5^ο κεφάλαιο, προτείνουμε τις βασικές συνθήκες για κάθε μια από τις κλασικές μεθόδους διαχωρισμού και επιπλέον, δείχνουμε τις ισχυρές ιδιότητες μοναδικότητας του μοντέλου PARAFAC, ειδικά για την περίπτωση του υπο-ορισμένου διαχωρισμού σημάτων (Underdetermined Blind source Separation).

Κεφάλαιο 1

Το Σύστημα Αναγνώρισης ομιλίας

1.1 Εισαγωγή

Για την αξιολόγηση της απόδοσης των μεθόδων του διαχωρισμού, χρησιμοποιούμε ένα ενιαίο σύστημα αναγνώρισης ομιλίας το οποίο ταξινομεί το φωνητικό περιεχόμενο από τις καθαρές (ανηχοϊκές) ομιλίες, τις μίξεις τους που λαμβάνονται από τα μικρόφωνα, καθώς και τα εκτιμώμενα σήματα ομιλίας που προκύπτουν μετά την εφαρμογή των μεθόδων διαχωρισμού. Με μέτρηση και σύγκριση του ποσοστού αναγνώρισης που επιτυγχάνεται με χρήση των σημάτων που έχουν υποστεί μείξη και των διαχωρισμένων σημάτων, μπορούν να προκύψουν συμπεράσματα που αφορούν την απόδοση των μεθόδων διαχωρισμού σε εφαρμογές αναγνώρισης ομιλίας. Στις επόμενες ενότητες παρουσιάζονται αναλυτικά όλα τα επιμέρους τμήματα του συστήματος αναγνώρισης ομιλίας.

1.2 Ποικιλία της Ομιλίας

Ο σημαντικότερος παράγοντας που επηρεάζει την απόδοση των συστημάτων ομιλίας είναι η μεγάλη ποικιλία της ομιλίας δηλαδή η διασπορά στο χώρο των παραμέτρων και στο πεδίο του χρόνου. Οι διαφορετικές πηγές της ποικιλίας της ομιλίας δημιουργούν τις περισσότερες δυσκολίες στην αναγνώριση ομιλίας. Κατά την παραγωγή ομιλίας, οι κινήσεις των διαφορετικών αρθρώσεων επικαλύπτονται στο χρόνο για διαδοχικά φωνητικά τμήματα που αλληλεπιδρούν μεταξύ τους. Σαν συνέπεια, η φωνητική οδός σε κάθε χρονική στιγμή επηρεάζεται από περισσότερα από ένα φωνητικά τμήματα. Αυτό το φαινόμενο ονομάζεται το φαινόμενο της συνάρθρωσης (*coarticulation*). Το φαινόμενο αυτό οφείλεται στο γεγονός ότι το ίδιο φώνημα μπορεί να παρουσιάζει ακουστικά χαρακτηριστικά, τα οποία εξαρτώνται από το περιβάλλον μέσα στο οποίο προφέρεται [1], [2]. Η παραγωγή ομιλίας επηρεάζεται εκτός από το φαινόμενο της συνάρθρωσης και από έναν μεγάλο αριθμό παραγόντων όπως οι τοπικοί διάλεκτοι, ο κοινωνικός παράγοντας, τα διαφορετικά χαρακτηριστικά της φωνητικής κοιλότητας του ομιλητή κ.α. Επίσης θα πρέπει να ληφθεί υπόψη και το περιβάλλον, τυχόν θόρυβοι, ηχώ καθώς και οι συνθήκες λειτουργίας του συστήματος αναγνώρισης ομιλίας, αφού προκαλούν τόσο ακουστικές όσο και φωνητικές μεταβολές. Συνεπώς, κάθε παραγόμενη ομιλία είναι μοναδική και αυτό έχει σαν αποτέλεσμα οι φυσικές διαδικασίες αποκωδικοποίησης του φωνητικού περιεχομένου των σημάτων ομιλίας που πραγματοποιούνται αυτόματα στα έμβια όντα να αποτελούν ένα πρόβλημα στην κατασκευή μιας αυτόματης μηχανής αναγνώρισης.

Παρόλο που αρκετοί τύποι ταξινομητών έχουν προταθεί για την αντιμετώπιση της διασποράς ομιλίας [3], [4], είναι αρκετά δύσκολο να υπάρξει σαφής διαχωρισμός μεταξύ τους λόγω της συνεχούς αλληλεπίδρασης τους.

Εδώ παρουσιάζονται δύο κύριες κατηγορίες ταξινομητών βασιζόμενοι στη διαφορά μεταξύ των παραλλαγών της ποικιλίας στον ίδιο ομιλητή (*intra – speaker*) και της μεταβλητότητας μεταξύ διαφορετικών ομιλητών (*inter – speaker*) :

1.2.1 Διαφορές στην ομιλία του ίδιου ομιλητή

Η ομιλία ενός ανθρώπου μπορεί να μεταβληθεί λόγω του διαφορετικού συγχρονισμού και του είδους των κινήσεων κατά την άρθρωση της ομιλίας. Γι' αυτό η ίδια ομιλία ή ακόμα και η ίδια λέξη, που παράγεται από τον ίδιο ομιλητή σε διαφορετικές καταστάσεις και κάτω από διαφορετικές συνθήκες, είναι συνήθως σημαντικά διαφορετική. Επίσης οι φυσικοί μηχανισμοί της ομιλίας δημιουργούν αλλαγές, που επιδρούν στην ηχώ της ρινικής κοιλότητας και στις δονήσεις των ακουστικών χορδών.

Λιγότερο φανερές είναι και οι αλλαγές στις βασικές συχνότητες και στον φωνητικό τύπο που προκαλούνται από αιτίες όπως η κούραση, το άγχος [5] ή ακόμα και η ηλικία [6]. Αλλαγές στην ένταση του ήχου που γίνονται συνειδητά, εμφανίζονται για παράδειγμα στην περίπτωση που φωνάζουμε σε μια άσχημη τηλεφωνική σύνδεση ή αυξάνουμε την ένταση της φωνής σε θορυβώδες περιβάλλον [8] ή σε μια φωνητική επικοινωνία που επηρεάζεται από την συναισθηματική μας κατάσταση: θυμός, χαρά κ.α.

1.2.2 Διαφορές στην ομιλία διαφορετικών ομιλητών

Οι διαφορές της ομιλίας ανάμεσα σε διαφορετικούς ομιλητές οφείλονται στην ανατομία των φωνητικών διόδων και των χορδών, στις διαλέκτους και στις τοπικές προφορές, στην κοινωνική τάξη, στο επίπεδο μόρφωσης στις ιδιοσυγκρασίες του λόγου κ.α. Είναι χρήσιμο να υπάρξει διάκριση μεταξύ των δύο κυριότερων μορφών των μεταβολών στην παραγωγή ήχου: τις φωνητικές και τις οργανικές [5].

Οι φωνητικές μεταβολές προέρχονται από διαφορές στον τρόπο με τον οποίο γίνεται η χρήση των οργάνων ομιλίας, ενώ οι οργανικές από τις διαφορές που οφείλονται στο ιδιαίτερο σχήμα των φωνητικών οργάνων. Η πιο ουσιαστική οργανική διαφορά εμφανίζεται ανάμεσα στην αντρική και στη γυναικεία φωνή [8]. Οι γυναίκες παρουσιάζουν υψηλότερο τόνο από τους άντρες, λόγω του διαφορετικού μήκους των φωνητικών τους χορδών (23-25mm στους άντρες , 15-17mm στις γυναίκες [9]).

Ο υψηλότερος τόνος μπορεί να επηρεάσει τη φασματική ανάλυση: στην πραγματικότητα ο τόνος εκτιμάται με λιγότερη ακρίβεια στις «υψηλές» φωνές. Γι' αυτό οι γυναικείες και παιδικές φωνές έχουν χρησιμοποιηθεί και εξεταστεί λιγότερο από τις αντρικές σε συστήματα αναγνώρισης ομιλίας, λόγω των μεγαλύτερων διαφορών και δυσκολιών στην εξαγωγή των ακουστικών τους παραμέτρων, και ειδικότερα στις βασικές συχνότητες γεγονός το οποίο μειώνει την αξιοπιστία της αναγνώρισης.

Το περιβάλλον επηρεάζει επίσης την παραγωγή, την καταληπτότητα και την ακουστική παρουσίαση της ομιλίας με έναν ιδιαίτερο τρόπο. Τα περιβαλλοντολογικά στοιχεία, χωρίζονται σε δύο κατηγορίες [Paoloni96]:

- Στατικά: ακουστική δωματίων, χρόνος ανάκλασης, συσκευές εγγραφής.

- Δυναμικά : θόρυβος, θέση μικροφώνου κ.α .

Γενικά υπάρχει μεγάλη δυσκολία στην μοντελοποίηση των στατικών και των δυναμικών περιβαλλοντολογικών παραγόντων λόγω της μεγάλης ποικιλομορφίας τους. Γι' αυτό και υπάρχει μεγάλη διαφορά στην απόδοση των συστημάτων που χρησιμοποιούνται σε περιβάλλοντα εργαστηρίου και αυτών που χρησιμοποιούνται σε πραγματικά περιβάλλοντα.

Τέλος πρέπει να λάβουμε υπόψη και την επίδραση του τοπικού και κοινωνικογλωσσικού περιβάλλοντος των ομιλητών. Οι άνθρωποι προφέρουν με διαφορετικό τρόπο τις λέξεις ανάλογα με τη γεωγραφική τους καταγωγή και ανάλογα με το γλωσσικό περιβάλλον των γονιών τους, το κοινωνικό τους επίπεδο και τη μόρφωσή τους. Ο τόνος και η τοπική προφορά διαφέρουν κυρίως στη δημιουργία των φωνημάτων και στα πρότυπα τονισμού καθώς οι διάλεκτοι διαφέρουν στη σύνταξη και στο λεξιλόγιο [10].

Ας δούμε τώρα πώς επιδρά η ποικιλομορφία της ομιλίας στην απόδοση των συστημάτων αναγνώρισης. Θα εξετάσουμε τρία θέματα: οι ομιλητές (εκπαιδευμένοι ή μη, στρεσαρισμένοι ή μη, κ.τ.λ), το περιβάλλον (επίπεδο θορύβου, εύρος ζώνης, παραμόρφωση) και το λεξιλόγιο (περιεχόμενο, μέγεθος, σύνταξη).

1. Εξάρτηση του ομιλητή. Δύο στρατηγικές μπορούν να ληφθούν υπόψη για το σχεδιασμό του συστήματος αναγνώρισης ομιλίας: Το σύστημα μπορεί να είναι είτε εξαρτημένο από τον ομιλητή ή ανεξάρτητο. Στην πρώτη περίπτωση το σύστημα είναι σχεδιασμένο να αναγνωρίζει την ομιλία ενός μόνο ομιλητή.

Στη δεύτερη περίπτωση το σύστημα είναι ικανό να αναγνωρίσει την ομιλία πολλών διαφορετικών ομιλητών. Ένα τέτοιο σύστημα απαιτεί μεγαλύτερη εκπαίδευση, αφού πρέπει να αντιμετωπίσει την μεγάλη μεταβλητότητα των διαφορετικών ομιλητών.

2. Η προσαρμογή του ομιλητή. Το σύστημα μπορεί να προσαρμοστεί στα χαρακτηριστικά γνωρίσματα των ομιλητών, βελτιώνοντας την αξιοπιστία της ταξινόμησης με τη χρήση. Η εκπαίδευση του επίσης μπορεί να προσαρμόζεται και ενώ το σύστημα βρίσκεται σε χρήση.

3. Παράμετροι ομιλίας. Τα δεδομένα της ομιλίας μέσα σε μια βάση δεδομένων μπορεί να είναι οτιδήποτε, από απομονωμένους ήχους μέχρι συνεχή ομιλία. Στη δεύτερη περίπτωση οι επιδράσεις του φαινομένου της συνάρθρωσης επηρεάζουν περισσότερο την απόδοση των συστημάτων αναγνώρισης.

4. Ρυθμός ομιλίας. Ο ρυθμός της προφοράς των λέξεων ποικίλει από ομιλητή σε ομιλητή καθώς επίσης και για τον ίδιο ομιλητή αφού εξαρτάται από τη φυσική και ψυχολογική κατάσταση του. Γι' αυτό η ομιλία μπορεί να είναι αργή, κανονική και γρήγορη. Ο ρυθμός της συνήθως μετράται από το μέσο αριθμό φωνημάτων που εμφανίζονται μέσα σε κάθε πρόταση.

5. Έξτρα ήχοι. Κατά την διάρκεια της ομιλίας παρουσιάζονται αρκετές φορές και ήχοι όπως για παράδειγμα βήχας, φτέρνισμα κλπ. Αυτές οι περιπτώσεις μπορούν να αντιμετωπισθούν σαν ιδιαίτερα τμήματα ομιλίας ή να αντιμετωπιστούν στο γλωσσικό επίπεδο.

6. Μέγεθος λεξιλογίου. Ο αριθμός των λέξεων, που μπορεί ένα σύστημα να αναγνωρίσει μπορεί να μεταβάλλεται από μερικές δεκάδες σε πολλές εκατοντάδες για συστήματα μεγάλου λεξιλογίου.

7. Δεδομένα εκπαίδευσης. Η διαδικασία της εκπαίδευσης είναι ένα κρίσιμο σημείο που περιέχει διαφορετικούς τύπους δεδομένων και πηγές πληροφορίας. Τα δεδομένα μπορεί να είναι είτε ηχογραφήσεις απομονωμένων λέξεων είτε φράσεις που μπορεί να περιλαμβάνουν και έξτρα ήχους.

Ο Πίνακας περιέχει τα χαρακτηριστικά, που πρέπει να ληφθούν υπόψη στο σχεδιασμό ενός συστήματος αναγνώρισης ομιλίας.

Περιβάλλον	Είδος θορύβου, λόγος σήματος προς θόρυβο
Διάταξη μετατροπής	Μικρόφωνο, τηλέφωνο
Κανάλι	Εύρος ζώνης συχνότητας, παραμόρφωση, ηχώ
Ομιλητές	Εξαρτημένου / Ανεξάρτητου ομιλητή, ηλικία, φύλο, κατάσταση
Είδος ομιλίας	Τόνος φωνής, ταχύτητα ομιλίας, διάλεκτος
Λεξιλόγιο	Χαρακτηριστικά των διαθέσιμων δεδομένων εκπαίδευσης, ειδικό ή γενικό λεξιλόγιο

Πίνακας Παράγοντες που πρέπει να ληφθούν υπόψη κατά την υλοποίηση συστήματος αναγνώρισης ομιλίας

1.3 Ακουστική και ομιλία

1.3.1 Ακουστική δωματίου

Η μετάδοση του ήχου από ένα σημείο ενός δωματίου σε ένα άλλο μπορεί να περιγραφεί με την κρουστική απόκριση του δωματίου μεταξύ των δύο σημείων θεωρώντας γραμμικό και χρονικά αμετάβλητο σύστημα. Η κρουστική απόκριση είναι μια συνάρτηση της πίεσης του αέρα που φτάνει σε ένα σημείο του δωματίου και προκαλείται από ένα ακουστικό παλμό που εκπέμπεται από ένα άλλο σημείο του δωματίου.

Χρησιμοποιώντας την υπόθεση της γραμμικής κυματικής υπέρθεσης στο σημείο του δέκτη, οι κρουστικές αποκρίσεις συνίστανται από τον συνδυασμό του απ' ευθείας ακουστικού μονοπατιού και των πολλαπλών ανακλάσεων του ήχου στους τοίχους του δωματίου. Το σήμα στον δέκτη είναι η υπέρθεση του σήματος της πηγής με διαφορετικές καθυστερήσεις και πλάτος

λόγω των διαφορετικών χρόνων διάδοσης και των διαφορετικών βαθμών απορρόφησης που υφίστανται από τους τοίχους, ή τα υπόλοιπα υλικά που υπάρχουν στο δωμάτιο. Ο βαθμός εξασθένησης του ανακλώμενου σήματος εξαρτάται κυρίως από το είδος του υλικού, την συχνότητα του ήχου και την γωνία της πρόσπτωσης. Τυπικά, οι πρώτες ανακλάσεις του ήχου μπορούν να παρατηρηθούν οπτικά από την κρουστική απόκριση. Όταν όμως ο αριθμός των ανακλάσεων γίνεται πολύ μεγάλος, η ουρά της κρουστικής απόκρισης παρουσιάζει πυκνότερη δομή λόγω της παρουσίας πολλαπλών υπερθέσεων. Ένα δωμάτιο μπορεί να χαρακτηριστεί από τον χρόνο αντήχησης ο οποίος ορίζεται σαν ο χρόνος που απαιτείται για να μειωθεί το επίπεδο της ακουστικής πίεσης του δωματίου κατά 60dB.

Ένας μικρός χρόνος αντήχησης (για παράδειγμα 100ms) αντιστοιχεί σε ένα «ξερό» δωμάτιο. Αντίθετα, ένας μεγάλος χρόνος αντήχησης αντιστοιχεί σε μια εκκλησία (6s) ή σε ένα συνεδριακό κέντρο (2-3s). Οι κρουστικές αποκρίσεις που φθίνουν απότομα αντιστοιχούν σε μικρούς χρόνους αντήχησης ενώ μεγάλες ουρές στην κρουστική απόκριση αντιστοιχούν σε μεγάλους χρόνους αντήχησης. Για ένα τυπικό γραφείο που παρουσιάζει χρόνο αντήχησης που προσεγγίζει τα 300ms μια αξιόπιστη κρουστική απόκριση θα πρέπει να αποτελείται από αρκετούς χιλιάδες συντελεστές ιδιαίτερα όταν η συχνότητα δειγματοληψίας είναι μεγάλη (για παράδειγμα 44.1kHz).

Παρόλο όμως που οι κρουστικές αποκρίσεις των δωματίων είναι θεωρητικά άπειρου μήκους, η χρήση μεγάλων φίλτρων στην επίλυση του προβλήματος του τυφλού διαχωρισμού δεν είναι κατά ανάγκη και η καλύτερη, αφού οι ουρές των κρουστικών αποκρίσεων έχουν πολύ μικρό πλάτος. Επίσης παρατηρείται αύξηση του σφάλματος εκτίμησης των συντελεστών του φίλτρου με αύξηση της τάξης των φίλτρων διαχωρισμού.

Σε πολλές πρακτικές εφαρμογές για τυπικά δωμάτια και συχνότητα δειγματοληψίας ίση με 16kHz η χρήση φίλτρων διαχωρισμού με περίπου χίλιους συντελεστές δίνει αρκετά αξιόπιστα αποτελέσματα.

1.3.2 Μικρόφωνα

Οι αποκρίσεις των μικροφώνων εξαρτώνται σε σημαντικό βαθμό από την συχνότητα δειγματοληψίας, την γωνία πρόσπτωσης του ηχητικού σήματος και την απόσταση από την ηχητική πηγή. Ο τύπος μικροφώνων που χρησιμοποιούνται στις ακουστικές εφαρμογές επηρεάζει σε μεγάλο βαθμό την αντήχηση και τον θόρυβο που καταγράφεται. Συγκεκριμένα, τα **κατευθυντικά** μικρόφωνα χρησιμοποιούνται για την καταγραφή ήχων που προέρχονται από συγκεκριμένες κατευθύνσεις. Στα κατευθυντικά μικρόφωνα οι ανακλάσεις και ο θόρυβος που υπάρχει στο δωμάτιο μπορούν να μειωθούν σε σημαντικό βαθμό, προκύπτοντας μικρότερος θόρυβος καταγραφής και αντηχήσεων. Θεωρητικά, τα **παντοκατευθυντικά** μικρόφωνα παρουσιάζουν την ίδια ευαισθησία για όλες τις γωνίες πρόσπτωσης του ηχητικού σήματος. Το σήμα που καταγράφεται περιέχει ανακλάσεις που προέρχονται από πολλές θέσεις του δωματίου, αυξάνοντας σημαντικά τις παραμορφώσεις λόγω αντήχησης και του θορύβου

του δωματίου. Τα **καρδιοειδή** μικρόφωνα παρουσιάζουν μια μέση κατευθυντικότητα που βρίσκεται μεταξύ αυτής των κατευθυντικών και των παντοκατευθυντικών μικροφώνων. Τυπικά είναι ευαίσθητα σε ήχους που προέρχονται από κωνική περιοχή ενώ παράλληλα παρουσιάζουν χαμηλότερη ευαισθησία καταγραφής σε όλες τις υπόλοιπες κατευθύνσεις. Για να μπορέσει να επιτευχθεί χαμηλός θόρυβος καταγραφής και μικρότερη αντήχηση, συνήθως χρησιμοποιούνται μικρόφωνα τα οποία παρουσιάζουν μικρή ενεργή περιοχή [11]. Ανάλογα με το είδος της εφαρμογής και την εκ των προτέρων γνώση σχετικά με τον αριθμό των ομιλητών και τις πιθανές θέσεις τους, η επιλογή του τύπου των μικροφώνων που θα χρησιμοποιηθούν μπορεί να καθοριστεί σε σημαντικό βαθμό. Η κατάλληλη επιλογή των σωστών μικροφώνων συμβάλει σημαντικά στην αξιόπιστη λειτουργία των αλγορίθμων διαχωρισμού.

Κεφάλαιο 2

Ανάλυση Ανεξάρτητων Συνιστωσών

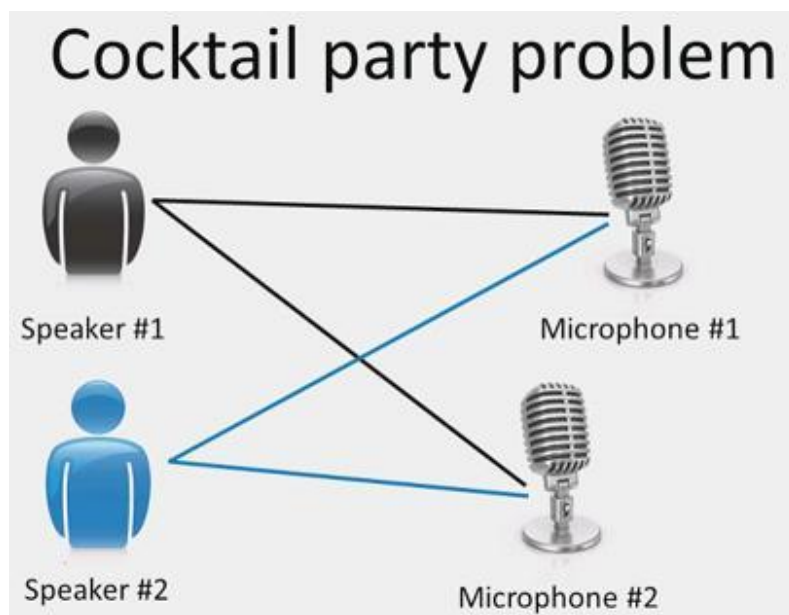
2.1 Σκοπός της μεθόδου

Υποθέτουμε ότι βρισκόμαστε σε ένα δωμάτιο όπου δύο άτομα μιλάνε ταυτόχρονα. Στο δωμάτιο υπάρχουν, επίσης, δύο μικρόφωνα, τα οποία τα τοποθετούμε σε συγκεκριμένη απόσταση μεταξύ τους. Τα μικρόφωνα μας δίνουν δύο ηχογραφημένα σήματα χρόνου τα οποία μπορούμε να τα συμβολίσουμε με $x_1(t)$ και $x_2(t)$, όπου x_1 και x_2 είναι τα πλάτη και t ο χρονικός δείκτης. Κάθε ένα από τα ηχογραφημένα σήματα το θεωρούμε σαν ένα σταθμισμένο σύνολο των σημάτων ομιλίας που εκπέμπονται από τα δύο μικρόφωνα, τα οποία τα συμβολίζουμε με $s_1(t)$ και $s_2(t)$. Αυτό μπορούμε να το εκφράσουμε σαν γραμμική εξίσωση:

$$x_1(t) = \alpha_{11}s_1 + \alpha_{12}s_2 \quad (2.1)$$

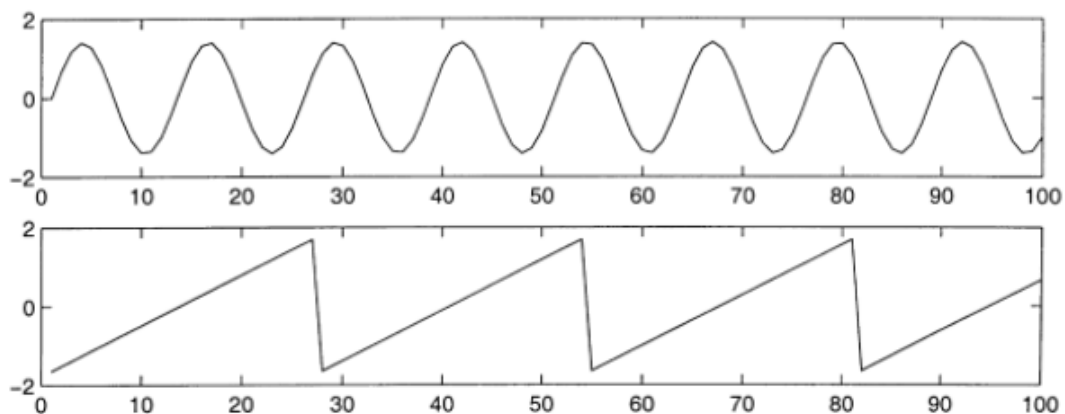
$$x_2(t) = \alpha_{21}s_1 + \alpha_{22}s_2 \quad (2.2)$$

όπου α_{11} , α_{12} , α_{21} και α_{22} είναι παράμετροι που εξαρτώνται από την απόσταση που έχουν τα μικρόφωνα από τους ομιλητές. Σ' αυτό το σημείο θα μας ήταν πολύ χρήσιμο αν μπορούσαμε να υπολογίσουμε τα δύο αρχικά σήματα ομιλίας $s_1(t)$ και $s_2(t)$, χρησιμοποιώντας μόνο τα καταγεγραμμένα σήματα $x_1(t)$ και $x_2(t)$. Ο φαινόμενο αυτό ονομάζεται «φαινόμενο του κοκτέιλ πάρτι» (cocktail party problem) (Εικόνα 2.1). Προς το παρόν παραλείπουμε χρονικές καθυστερήσεις ή άλλους παράγοντες από το απλοποιημένο σύστημα μίξης.

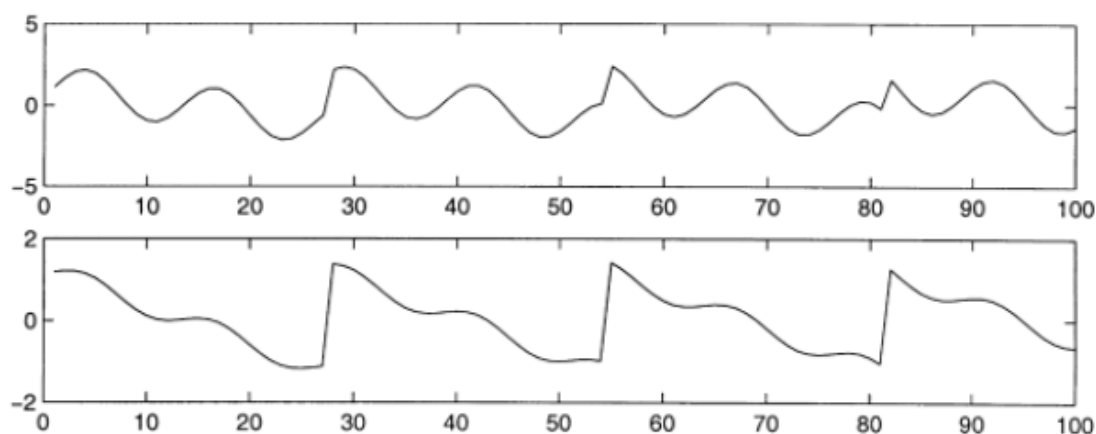


Εικόνα 2.1 Το φαινόμενο του κοκτέιλ πάρτι.

Αν θέλουμε να απεικονίσουμε γραφικά το πρόβλημα τότε θεωρούμε τις παρακάτω κυματομορφές (Εικόνα 2.2 και Εικόνα 2.3). Είναι σαφές ότι δεν χρησιμοποιούμε πραγματικά σήματα ομιλίας αλλά επαρκούν για το παράδειγμά μας.



Εικόνα 2.2 Τα αρχικά σήματα.

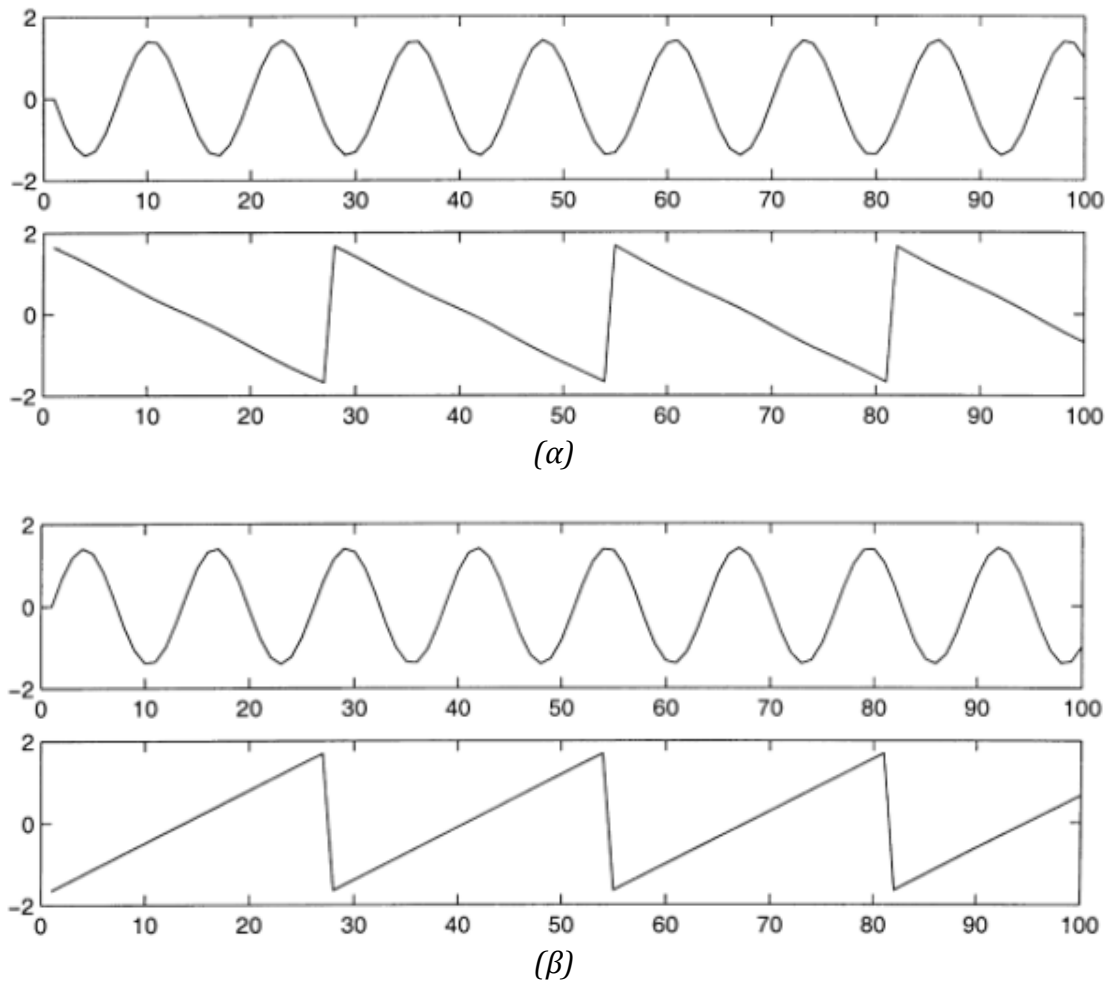


Εικόνα 2.3 Η μίξη των σημάτων πηγής της Εικόνας 2.1.

Τα αρχικά σήματα ομιλίας θα έμοιαζαν όπως αυτά στην Εικόνα 2.2 και τα σήματα μετά τη μίξη θα έμοιαζαν όπως αυτά στην Εικόνα 2.3. Το πρόβλημα που προκύπτει λοιπόν είναι να ανακτηθούν τα δεδομένα της Εικόνας 2.2 χρησιμοποιώντας μόνο τα δεδομένα από την Εικόνα 2.3.

Στην πραγματικότητα, αν γνωρίζαμε τις παραμέτρους α_{ij} , θα μπορούσαμε να λύσουμε την γραμμική εξίσωση (2.1) με κλασσικές μεθόδους. Το θέμα είναι, ωστόσο, ότι εφόσον δεν γνωρίζουμε το α_{ij} , το πρόβλημα γίνεται αρκετά πιο δύσκολο.

Μια προσέγγιση για τη λύση του προβλήματος θα ήταν να χρησιμοποιήσουμε κάποια πληροφορία για τις στατιστικές ιδιότητες των σημάτων $s_i(t)$ για να υπολογίσουμε το α_{ij} . Πράγματι, και μάλλον όχι αναμενόμενα, αποδεικνύεται ότι είναι αρκετό να θεωρήσουμε ότι τα $s_1(t)$ και $s_2(t)$, σε κάθε χρονική στιγμή t , είναι *στατιστικά ανεξάρτητα*. Η υπόθεση αυτή είναι αρκετά ρεαλιστική σε πολλές περιπτώσεις και δεν χρειάζεται να είναι ακριβώς αληθινή στην πράξη. Η τεχνική της Ανάλυσης Ανεξάρτητων Συνιστωσών (Independent Component Analysis or ICA), μπορεί να χρησιμοποιηθεί για να υπολογίσουμε το α_{ij} , βασιζόμενοι πάντα στην πληροφορία για την ανεξαρτησία τους, η οποία μας επιτρέπει να διαχωρίσουμε τα δύο αρχικά σήματα $s_1(t)$ και $s_2(t)$ από τα σήματα της μίξης τους, τα $\chi_1(t)$ και $\chi_2(t)$. Η Εικόνα 2.4 μας δίνει τα δύο σήματα που υπολογίζονται με τη μέθοδο της Ανάλυσης Ανεξάρτητων Συνιστωσών. Όπως φαίνεται, τα σήματα που προκύπτουν είναι πολύ κοντά στα αρχικά μας σήματα (παρατηρούμε ότι τα πρόσημά τους είναι ανεστραμμένα αλλά αυτό δεν έχει ιδιαίτερη σημασία).



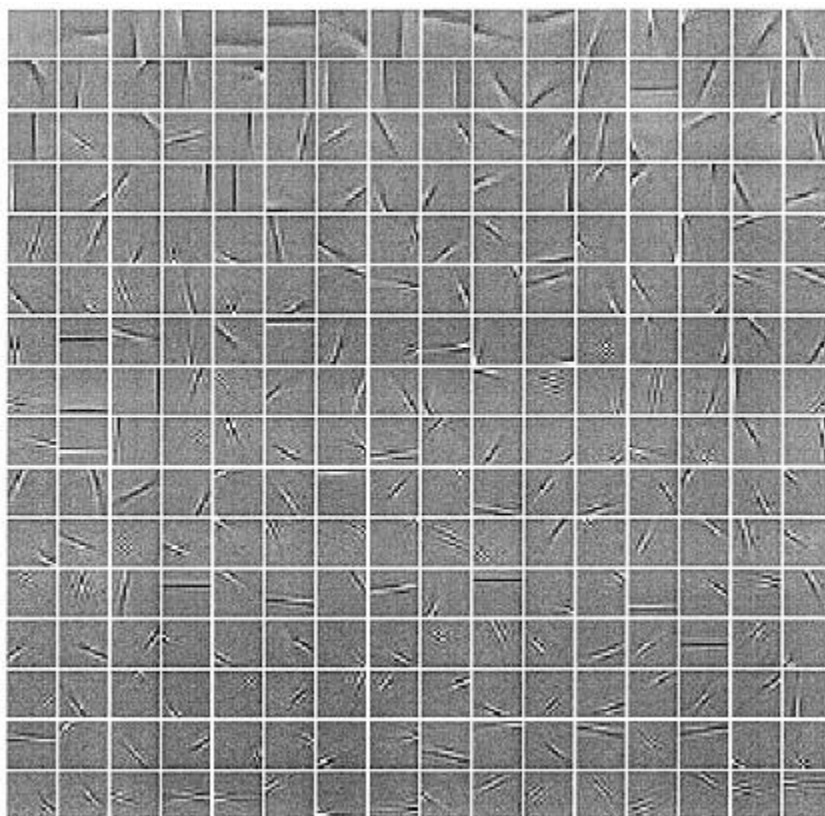
Εικόνα 2.4 Οι εκτιμήσεις (α) των αρχικών σημάτων πηγής (β) υπολογισμένες χρησιμοποιώντας μόνο τα παρατηρούμενα σήματα της Εικόνας 2.3.

Η ανάλυση ανεξάρτητων συνιστωσών αναπτύχθηκε αρχικά για να επιλυθούν προβλήματα που σχετίζονται με το πρόβλημα του κοκτέιλ πάρτι. Με την πρόσφατη αύξηση του ενδιαφέροντος για την ICA, έγινε σαφές ότι αυτή η μέθοδος έχει πολλές ακόμα ενδιαφέρουσες εφαρμογές. Θεωρούμε για παράδειγμα, ηλεκτρικές καταγραφές της εγκεφαλικής δραστηριότητας όπως δίνονται από ένα ηλεκτροεγκεφαλογράφημα (ΗΕΓ). Τα δεδομένα του ΗΕΓ περιλαμβάνουν καταγραφές ηλεκτρικού δυναμικού από πολλές διαφορετικές περιοχές του κρανίου. Αυτές οι καταγραφές προφανώς προέκυψαν από τη μίξη συγκεκριμένων συνιστωσών της εγκεφαλικής δραστηριότητας. Αυτή η κατάσταση μοιάζει αρκετά με την περίπτωση του κοκτέιλ πάρτι: Επιθυμούμε να βρούμε τις αρχικές συνιστώσες της εγκεφαλικής δραστηριότητας, αλλά μπορούμε να παρατηρήσουμε μόνο τη μίξη των συνιστωσών αυτών. Η μέθοδος ICA μπορεί να αποκαλύψει σημαντικές πληροφορίες για την εγκεφαλική δραστηριότητα δίνοντας μας πρόσβαση στις ανεξάρτητες συνιστώσες της.

Μία άλλη πολύ διαφορετική εφαρμογή της ICA είναι για την εξαγωγή

χαρακτηριστικών. Ένα βασικό πρόβλημα στην ψηφιακή επεξεργασία σήματος είναι να βρούμε κατάλληλες αναπαραστάσεις για εικόνα, ήχο ή κάποιο άλλο είδος δεδομένων για διεργασίες όπως συμπίεση ή αφαίρεση θορύβου. Η αναπαράσταση των δεδομένων συχνά βασίζεται σε (διακριτούς) γραμμικούς μετασχηματισμούς. Συνήθεις γραμμικοί μετασχηματισμοί που χρησιμοποιούνται ευρέως στην επεξεργασία εικόνας είναι ο μετασχηματισμός Fourier, Haar, μετασχηματισμός συνημίτονου κ.α.. Καθένας έχει τις δικές του χαρακτηριστικές ιδιότητες[12].

Θα ήταν ιδιαίτερα χρήσιμο να υπολογίσουμε τον γραμμικό μετασχηματισμό από τα ίδια τα δεδομένα, μιας και στην περίπτωση αυτή ο μετασχηματισμός θα ταίριαζε ιδανικά στα δεδομένα τα οποία επεξεργαζόμαστε. Η Εικόνα 2.5 δείχνει τις συναρτήσεις βάσης που αποκτούμε από την ICA από τμήματα φυσικών εικόνων. Κάθε παράθυρο εικόνας (image window) στο σύνολο των εικόνων θα μπορούσε να είναι μια υπέρθεση από αυτά τα παράθυρα έτσι ώστε ο συντελεστής στην υπέρθεση να είναι ανεξάρτητος. Η εξαγωγή χαρακτηριστικών χρησιμοποιώντας την ICA θα εξηγηθεί λεπτομερώς σε επόμενο κεφάλαιο.



Εικόνα 2.5 Συναρτήσεις βάσης της ICA από φυσικές εικόνες. Το μέγεθος του παραθύρου εισαγωγής ήταν 16×16 εικονοστοιχεία. Αυτές οι συναρτήσεις βάσεις μπορούν να θεωρηθούν ανεξάρτητα χαρακτηριστικά των εικόνων.

Όλες οι εφαρμογές που περιγράφονται παραπάνω μπορούν να διατυπωθούν σε ένα ενιαίο μαθηματικό πλαίσιο, αυτό της ICA. Αυτό είναι μία γενική μέθοδος της επεξεργασίας σήματος και της ανάλυσης δεδομένων.

2.2 Ανάλυση Ανεξάρτητων Συνιστωσών

2.2.1 Ορισμός της ICA

Για να ορίσουμε αυστηρά την ICA [13], μπορούμε να χρησιμοποιήσουμε ένα στατιστικό μοντέλο «λανθανουσών μεταβλητών». Θεωρούμε ότι παρατηρούμε n γραμμικούς συνδυασμούς $\chi_1, \dots, \chi_n$ n ανεξάρτητων συνιστωσών.

$$\chi_j = \alpha_{j1}s_1 + \alpha_{j2}s_2 + \dots + \alpha_{jn}s_n, \forall j \quad (2.3)$$

Έχουμε παραλείψει τη μεταβλητή του χρόνου t , καθώς στο μοντέλο ICA υποθέτουμε ότι κάθε συνδυασμός χ_j όπως και κάθε ανεξάρτητη συνιστώσα s_k είναι μία τυχαία μεταβλητή και όχι ένα συγκεκριμένο σήμα χρόνου. Οι παρατηρούμενες μεταβλητές $\chi_j(t)$, π.χ. τα σήματα των μικροφώνων στο πρόβλημα του κοκτέιλ πάρτι, είναι ένα δείγμα αυτής της τυχαίας μεταβλητής. Χωρίς βλάβη της γενικότητας, μπορούμε να υποθέσουμε ότι και οι συνδυασμένες μεταβλητές και οι ανεξάρτητες συνιστώσες έχουν μηδενική μέση τιμή. Αν αυτό δεν ισχύει, τότε οι παρατηρούμενες μεταβλητές χ_i μπορούν πάντα να αποκτήσουν μηδενική μέση τιμή αφαιρώντας το δειγματικό μέσο, και έτσι προκύπτει το μοντέλο μηδενικού μέσου όρου.

Είναι βολικό να χρησιμοποιήσουμε συμβολισμό διανύσματος-πίνακα αντί για τα αθροίσματα που χρησιμοποιήσαμε στην εξίσωση 2.3. Ας συμβολίσουμε με $\mathbf{x}$ το τυχαίο διάνυσμα του οποίου τα στοιχεία είναι οι συνδυασμοί $\chi_1, \dots, \chi_n$, και με $\mathbf{s}$ το τυχαίο διάνυσμα με στοιχεία τα $s_1, \dots, s_n$. Επίσης, συμβολίζουμε με $\mathbf{A}$ τον πίνακα με στοιχεία α_{ij} . Γενικά, τα έντονα (Bold) πεζά γράμματα συμβολίζουν διανύσματα και τα έντονα κεφαλαία γράμματα συμβολίζουν πίνακες. Όλα τα διανύσματα θεωρούνται διανύσματα στήλες και συνεπώς ο $\mathbf{x}^T$, ή ο ανάστροφος του $\mathbf{x}$, είναι διάνυσμα γραμμή. Χρησιμοποιώντας αυτόν τον διανυσματικό πλέον πίνακα, το παραπάνω μοντέλο μίξης μπορεί γραφεί ως εξής:

$$\mathbf{x} = \mathbf{A}\mathbf{s} \quad (2.4)$$

Μερικές φορές χρειαζόμαστε τις στήλες του πίνακα $\mathbf{A}$ · συμβολίζοντάς τες με $\mathbf{a}_j$ το μοντέλο μπορεί επιπλέον να γραφεί ως εξής:

$$\mathbf{x} = \sum_{i=1}^n \mathbf{a}_i s_i \quad (2.5)$$

Το στατιστικό μοντέλο στην εξίσωση 2.4 καλείται *ανάλυση των ανεξάρτητων συνιστωσών ή μοντέλο ICA*. Το μοντέλο ICA είναι ένα παραγωγικό μοντέλο, το οποίο σημαίνει ότι περιγράφει πώς τα παρατηρούμενα δεδομένα αναπαράγονται μέσα από μια διαδικασία μίξης των συνιστωσών s_i . Οι ανεξάρτητες συνιστώσες είναι λανθάνουσες μεταβλητές, γεγονός που σημαίνει ότι δεν μπορούν να παρατηρηθούν άμεσα. Επιπλέον, ο πίνακας μίξης θεωρείται ότι είναι άγνωστος. Αυτό που παρατηρούμε είναι το τυχαίο διάνυσμα $\mathbf{x}$, και πρέπει να υπολογίσουμε και το $\mathbf{A}$ και το $\mathbf{s}$ χρησιμοποιώντας το $\mathbf{x}$. Αυτό πρέπει να γίνει κάνοντας όσο το δυνατόν πιο γενικές υποθέσεις. Το σημείο εκκίνησης για την ICA είναι η πολύ απλή υπόθεση ότι οι συνιστώσες s_i είναι *στατιστικά ανεξάρτητες*. Πρέπει επίσης να υποθέσουμε ότι η ανεξάρτητη συνιστώσα πρέπει να έχει *μη-Γκαουσιανές* κατανομές. Ωστόσο, στο βασικό μοντέλο δεν θεωρούμε αυτές τις κατανομές γνωστές (αν είναι γνωστές, το πρόβλημα θεωρείται απλοποιημένο). Για ευκολία, υποθέτουμε ότι ο άγνωστος πίνακας μίξης είναι τετραγωνικός, αλλά αυτή η υπόθεση μπορεί ορισμένες φορές να μην είναι τόσο αυστηρή. Αφού υπολογίσουμε τον πίνακα $\mathbf{A}$, μπορούμε να υπολογίσουμε τον αντίστροφό του, έστω $\mathbf{W}$, και να καταλήξουμε εύκολα στην ανεξάρτητη συνιστώσα μέσω της σχέσης:

$$\mathbf{s} = \mathbf{W}\mathbf{x} \quad (2.6)$$

Η ICA είναι μία μέθοδος στενά συνδεδεμένη με τη μέθοδο του *Τυφλού Διαχωρισμού Πηγαίων Σημάτων (Blind Source Separation or BSS)*. Ο όρος «Πηγαίων» (“Source”) αναφέρεται στο αρχικό μας σήμα π.χ. ανεξάρτητη συνιστώσα, όπως το μικρόφωνο σε ένα πρόβλημα κοκτέιλ πάρτι. Ο όρος «Τυφλός» (“Blind”) σημαίνει ότι γνωρίζουμε πολύ λίγα, αν όχι τίποτα, για τον πίνακα μίξης, και κάνουμε κάποιες υποθέσεις για τις σήματα των πηγών. Η ICA είναι μία μέθοδος, ίσως η πιο ευρέως χρησιμοποιούμενη, για την εκτέλεση του τυφλού διαχωρισμού πηγαίων σημάτων.

Σε πολλές εφαρμογές, θα ήταν πιο ρεαλιστικό να υποθέσουμε ότι υπάρχει και θόρυβος στις μετρήσεις μας [14],[15], το οποίο θα σήμαινε την πρόσθεση ενός νέου όρου στο μοντέλο ο οποίος θα συμβολίζει τον θόρυβο. Για λόγους όμως απλούστευσης παραλείπουμε όλους τους όρους για το θόρυβο, αφού ο υπολογισμός του χωρίς θόρυβο μοντέλου είναι ήδη αρκετά δύσκολος και φαίνεται να είναι επαρκής για πολλές εφαρμογές.

2.2.2 Ασάφειες της ICA

Στο μοντέλο της ICA στην εξίσωση 2.4, εύκολα μπορούμε να διακρίνουμε τις ακόλουθες ασάφειες:

1. Δεν μπορούμε να ορίσουμε επακριβώς τις μεταβλητές των ανεξάρτητων συνιστωσών.

Ο λόγος είναι ότι, εφόσον το $\mathbf{x}$ και το $\mathbf{A}$ είναι άγνωστα, κάθε βαθμωτός πολλαπλασιαστής σε μία από τις πηγές s_i θα μπορούσε πάντα να απαλείφεται διαιρώντας την αντίστοιχη στήλη $\mathbf{a}_{ij}$ του $\mathbf{A}$ με το ίδιο βαθμωτό μέγεθος (βλέπε εξίσωση 2.5). Συνεπώς, μπορούμε κατά κάποιο τρόπο να καθορίσουμε τα μεγέθη των ανεξάρτητων συνιστωσών· εφόσον είναι τυχαίες μεταβλητές, ο πιο λογικός τρόπος είναι να το κάνουμε αυτό είναι να υποθέσουμε ότι η καθεμία έχει μοναδιαία διασπορά: $E\{s_i^2\}=1$. Έπειτα ο πίνακας $\mathbf{A}$ προσαρμόζεται στις μεθόδους επίλυσης της ICA έτσι ώστε να λάβουμε υπ' όψιν αυτόν τον περιορισμό. Να σημειωθεί ότι υπάρχει ακόμα ασάφεια για το πρόσημο: θα μπορούσαμε να πολλαπλασιάσουμε την ανεξάρτητη συνιστώσα με -1 χωρίς να επηρεαστεί το μοντέλο. Αυτή η ασάφεια, ευτυχώς, δεν είναι σημαντική σε πολλές εφαρμογές.

2. Δεν μπορούμε να καθορίσουμε τη σειρά των ανεξάρτητων συνιστωσών.

Ο λόγος είναι ότι, εφόσον τα $\mathbf{x}$ και $\mathbf{A}$ εξακολουθούν να είναι άγνωστα, μπορούμε ελεύθερα να αλλάξουμε τη σειρά των όρων στην εξίσωση 2.5, και να καλέσουμε οποιαδήποτε ανεξάρτητη συνιστώσα σαν πρώτη. Τυπικά, ένας πίνακας αντιμετάθεσης $\mathbf{P}$ και ο αντίστροφός του μπορεί να αντικατασταθεί στο μοντέλο ώστε να δώσουν $\mathbf{x}=\mathbf{A}\mathbf{P}^{-1}\mathbf{P}\mathbf{s}$. Τα στοιχεία του $\mathbf{P}\mathbf{s}$ είναι οι αρχικές ανεξάρτητες μεταβλητές s_j , αλλά με άλλη σειρά. Ο πίνακας $\mathbf{A}\mathbf{P}^{-1}$ είναι ένας καινούριος άγνωστος πίνακας μίξης, ο οποίος θα λυθεί με τους αλγόριθμους της ICA.

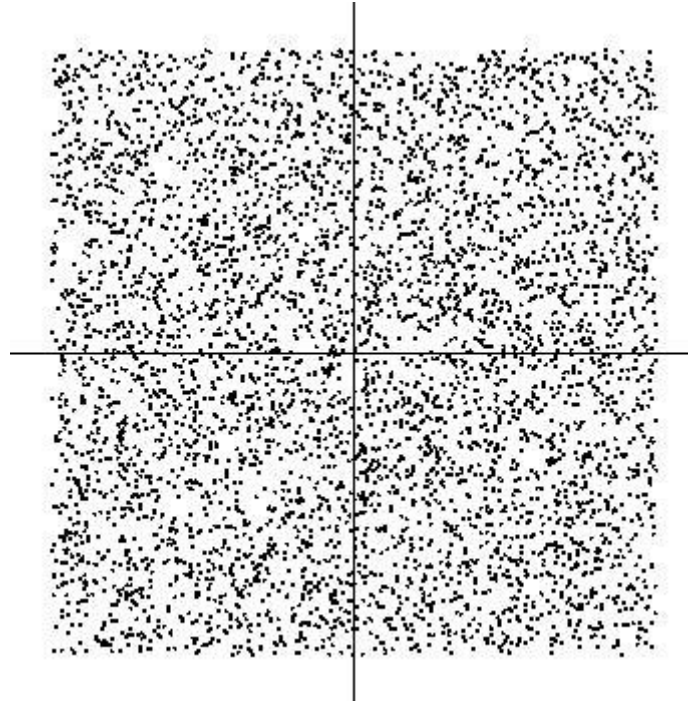
2.2.3 Απεικόνιση της ICA

Για να απεικονίσουμε το μοντέλο της ICA με στατιστικούς όρους, θεωρούμε δύο ανεξάρτητες συνιστώσες που έχουν την ακόλουθη ομοιόμορφη κατανομή:

$$p(s_i) = \begin{cases} \frac{1}{2\sqrt{3}} & \text{αν } |s_i| \leq \sqrt{3} \\ 0 & \text{αλλιού} \end{cases} \quad (2.7)$$

Το πεδίο των τιμών για αυτή την ομοιόμορφη κατανομή επιλέχτηκε έτσι ώστε να προκύψει μέση τιμή μηδέν και διασπορά ίση με ένα, όπως αναφέραμε στην προηγούμενη ενότητα. Η από κοινού πυκνότητα πιθανότητας των s_1 και s_2 είναι

τώρα η τετραγωνική ομοιόμορφη κατανομή. Αυτό συνεπάγεται από τον βασικό ορισμό ότι η από κοινού πυκνότητα πιθανότητας δύο ανεξάρτητων μεταβλητών είναι το γινόμενο των περιθωρίων πυκνοτήτων πιθανότητάς τους (βλέπε εξίσωση 2.10). Συνεπώς πρέπει να υπολογιστεί το γινόμενο. Η από κοινού πυκνότητα πιθανότητας απεικονίζεται στην Εικόνα 6 όπου φαίνονται δεδομένα τυχαία επιλεγμένα από αυτή την κατανομή.

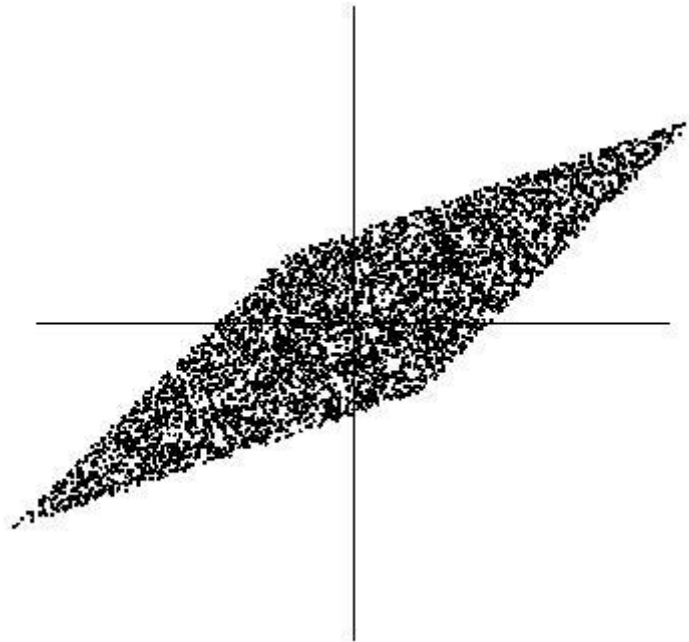


Εικόνα 2.6 Η από κοινού συνάρτηση πυκνότητας πιθανότητας των ανεξάρτητων μεταβλητών s_1 και s_2 με ομοιόμορφες κατανομές. Οριζόντιος άξονας: s_1 , κατακόρυφος άξονας: s_2 .

Συνδυάζοντας αυτές τις δύο ανεξάρτητες συνιστώσες παίρνουμε τον ακόλουθο πίνακα μίξης:

$$\mathbf{A}_0 = \begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix} \quad (2.8)$$

Αυτό μας δίνει δύο συνδυασμένες μεταβλητές, χ_1 και χ_2 . Εύκολα υπολογίζεται ότι τα συνδυασμένα δεδομένα έχουν ομοιόμορφη κατανομή ένα παραλληλόγραμμο, όπως φαίνεται στην Εικόνα 2.7.



Εικόνα 2.7 Η από κοινού συνάρτηση πυκνότητας πιθανότητας των παρατηρούμενων συνδυασμών χ_1 και χ_2 . Οριζόντιος άξονας: s_1 , κατακόρυφος άξονας: s_2 .

Επισημαίνεται ότι οι μεταβλητές χ_1 και χ_2 δεν είναι πλέον ανεξάρτητες· ένας εύκολος τρόπος να το δούμε αυτό είναι να θεωρήσουμε, εφόσον είναι δυνατόν να προβλέψουμε την τιμή ενός από αυτά, έστω της χ_2 , από την τιμή της άλλης. Προφανώς αν η χ_1 πάρει μία από τις μέγιστες ή ελάχιστες τιμές, τότε αυτό καθορίζει πλήρως την τιμή του χ_2 . Συνεπώς δεν είναι ανεξάρτητες. (Για μεταβλητές s_1 και s_2 η κατάσταση είναι διαφορετική: από την Εικόνα 2.6 φαίνεται ότι η γνωστή τιμή του s_1 δεν μας βοηθάει καθόλου να υποθέσουμε ποια μπορεί να είναι η τιμή του s_2).

Το πρόβλημα του υπολογισμού των δεδομένων του μοντέλου της ICA είναι πλέον να υπολογίσουμε τον πίνακα $\mathbf{A}_0$ χρησιμοποιώντας μόνο την πληροφορία που περιέχεται στους συνδυασμούς των χ_1 και χ_2 . Πράγματι, από την Εικόνα 2.7 μπορούμε να διακρίνουμε έναν διαισθητικό τρόπο για να υπολογίσουμε τον $\mathbf{A}$: οι ακμές του παραλληλόγραμμου έχουν την διεύθυνση των στηλών του $\mathbf{A}$. Αυτό σημαίνει ότι θα μπορούσαμε, σε γενικές γραμμές, να υπολογίσουμε το μοντέλο της ICA αφού πρώτα υπολογίσουμε την από κοινού πυκνότητα πιθανότητας των χ_1 και χ_2 , και στη συνέχεια να εντοπίσουμε τις ακμές. Έτσι, το πρόβλημα φαίνεται να έχει λύση.

Στην πραγματικότητα, ωστόσο, αυτό θα ήταν μια όχι και τόσο καλή λύση διότι δουλεύει μόνο με μεταβλητές που έχουν ακριβώς ομοιόμορφες

κατανομές. Επιπλέον, θα ήταν αρκετά περίπλοκο όσον αφορά στους υπολογισμούς. Αυτό που χρειαζόμαστε είναι μία μέθοδος που δουλεύει για κάθε κατανομή των ανεξάρτητων συνιστωσών και δουλεύει γρήγορα και αξιόπιστα.

2.3 Περί ανεξαρτησίας

2.3.1 Ορισμός και βασικές ιδιότητες

Για να ορίσουμε την έννοια της ανεξαρτησίας, θεωρούμε δύο βαθμωτές τυχαίες μεταβλητές y_1 και y_2 . Βασικά, οι δύο αυτές μεταβλητές θεωρούνται ανεξάρτητες αν η πληροφορία στην τιμή της y_1 δεν δίνει καμία πληροφορία στην τιμή της y_2 , και αντιστρόφως. Πιο πάνω, επισημάναμε ότι αυτή είναι η περίπτωση με τις μεταβλητές s_1, s_2 αλλά χωρίς τους συνδυασμούς των μεταβλητών x_1 και x_2 .

Με την αυστηρή ερμηνεία του όρου, η ανεξαρτησία μπορεί να οριστεί από τις πυκνότητες πιθανότητας. Συμβολίζουμε με $p_1(y_1)$ την από κοινού συνάρτηση πυκνότητας πιθανότητας (probability density function or pdf) της y_1 , π.χ. η συνάρτηση πυκνότητας πιθανότητας μόνο της y_1 είναι:

$$p_1(y_1) \int p_1(y_1, y_2) dy_2 \quad (2.9)$$

Και όμοια για την y_2 . Στη συνέχεια θεωρούμε ότι τα y_1 και y_2 είναι ανεξάρτητα αν και μόνον αν η από κοινού πυκνότητα πιθανότητας μπορεί να γραφεί με τον ακόλουθο τρόπο:

$$p_1(y_1, y_2) = p_1(y_1)p_2(y_2) \quad (2.10)$$

Αυτός ο ορισμός μπορεί να επεκταθεί φυσικά για κάθε αριθμό n τυχαίων μεταβλητών, όπου στην περίπτωση αυτή η από κοινού πυκνότητα πιθανότητας πρέπει να είναι ένα γινόμενο n όρων.

Ο ορισμός μπορεί να χρησιμοποιηθεί για να εξάγου μια πολύ σημαντική ιδιότητα των ανεξάρτητων τυχαίων μεταβλητών. Έστω ότι δίνονται οι συναρτήσεις h_1 και h_2 , πάντα ισχύει:

$$E\{h_1(y_1)h_2(y_2)\} = E\{h_1(y_1)\}E\{h_2(y_2)\} \quad (2.11)$$

Αυτό μπορεί να αποδειχθεί ως εξής:

$$\begin{aligned} E\{h_1(y_1)h_2(y_2)\} &= \iint h_1(y_1) h_2(y_2) p(y_1, y_2) dy_1 dy_2 \\ &= \iint h_1(y_1) p_1(y_1) h_2(y_2) p_2(y_2) dy_1 dy_2 \end{aligned}$$

$$\begin{aligned}
 &= \int h_1(y_1) p_1(y_1) dy_1 \int h_2(y_2) p_2(y_2) dy_2 \\
 &= E\{h_1(y_1)\} E\{h_2(y_2)\}.
 \end{aligned} \tag{2.12}$$

2.3.2 Οι ασυσχέτιστες μεταβλητές είναι μόνο μερικώς ανεξάρτητες

Μία λιγότερο αυστηρή μορφή ανεξαρτησίας είναι η μη-συσχέτιση. Δύο τυχαίες μεταβλητές y_1 και y_2 είναι ασυσχέτιστες, αν η συνδιασπορά τους είναι μηδέν:

$$E\{y_1 y_2\} - E\{y_1\} E\{y_2\} = 0 \tag{2.13}$$

Αν οι μεταβλητές είναι ανεξάρτητες, είναι και ασυσχέτιστες, το οποίο προκύπτει άμεσα από την εξίσωση 2.11, θεωρώντας $h_1(y_1) = y_1$ και $h_2(y_2) = y_2$.

Αντίθετα, η μη-συσχέτιση δε σημαίνει απαραίτητα και ανεξαρτησία. Για παράδειγμα, υποθέτουμε ότι (y_1, y_2) είναι διακριτές μεταβλητές και ακολουθούν τέτοια κατανομή ώστε με πιθανότητα $1/4$ να παίρνουν οποιοδήποτε από τα ακόλουθα ζεύγη τιμών: $(0,1)$, $(0,-1)$, $(1,0)$, $(-1,0)$. Τότε τα y_1 και y_2 είναι ασυσχέτιστα. Όμως,

$$E\{y_1^2 y_2^2\} = 0 \neq \frac{1}{4} = E\{y_1^2\} E\{y_2^2\} \tag{2.14}$$

πράγμα που σημαίνει ότι η εξίσωση 2.11 παραβιάζεται, και συνεπώς οι μεταβλητές δεν μπορούν να είναι αμετάβλητες.

Εφόσον η ανεξαρτησία σημαίνει και τη μη-συσχέτιση, πολλές μέθοδοι ICA εξαναγκάζουν την διαδικασία υπολογισμού, έτσι ώστε να μας δίνει πάντα ασυσχέτιστες τιμές των ανεξάρτητων συνιστωσών. Αυτό μειώνει τον αριθμό των παραμέτρων και απλοποιεί το πρόβλημα.

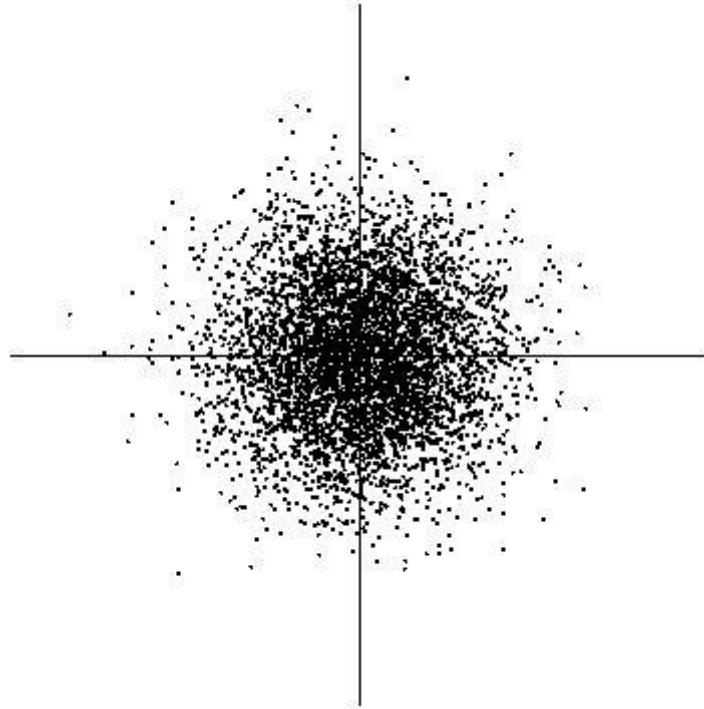
2.3.3 Γιατί οι Γκαουσιανές μεταβλητές είναι απαγορευμένες

Ο βασικός περιορισμός της ICA είναι ότι οι ανεξάρτητες συνιστώσες πρέπει να είναι μη-Γκαουσιανές έτσι ώστε να είναι δυνατή η εφαρμογή της.

Για να κατανοήσουμε γιατί οι Γκαουσιανές μεταβλητές κάνουν δυνατή την εφαρμογή της ICA, υποθέτουμε ότι ο πίνακας μίξης είναι ορθογώνιος και τα s_i είναι Γκαουσιανές μεταβλητές. Τότε και τα x_1 και x_2 είναι Γκαουσιανές μεταβλητές, και με μοναδιαία διασπορά. Η από κοινού πυκνότητα πιθανότητάς τους δίνεται από τη σχέση:

$$p(x_1, x_2) = \frac{1}{2\pi} \exp\left(-\frac{x_1^2 + x_2^2}{2}\right) \tag{2.15}$$

Αυτή η κατανομή παριστάνεται στην Εικόνα 2.8 όπου φαίνεται ότι η πυκνότητα πιθανότητας είναι απόλυτα συμμετρική. Επομένως, δεν περιέχει καμία πληροφορία για τις κατευθύνσεις των στηλών του πίνακα μίξης **A**. Αυτός είναι ο λόγος που ο πίνακας **A** δεν μπορεί να υπολογιστεί.



Εικόνα 2.8 Η πολυμεταβλητή κατανομή δύο ανεξάρτητων Γκαουσιανών μεταβλητών.

Με μια πιο διεξοδική ανάλυση, μπορεί να αποδειχθεί ότι η κατανομή οποιουδήποτε ορθογώνιου μετασχηματισμού των Γκαουσιανών μεταβλητών (χ_1, χ_2) έχει ακριβώς την ίδια κατανομή με τις (χ_1, χ_2) και ότι τα χ_1 και χ_2 είναι ανεξάρτητες. Στην πραγματικότητα, αν μόνο μία από τις ανεξάρτητες μεταβλητές είναι Γκαουσιανή, το μοντέλο της ICA μπορεί και πάλι να εφαρμοστεί.

2.4 Παραδοχές υπολογισμού της ICA

2.4.1 «Η μη-Γκαουσιανή είναι ανεξάρτητη»

Σύμφωνα με την παραπάνω ανάλυση, το κλειδί για να υπολογίσουμε το μοντέλο της ICA είναι το γεγονός ότι είναι μη-Γκαουσιανό. Πράγματι, χωρίς το γεγονός αυτό ο υπολογισμός θα ήταν αδύνατος. Αυτή είναι ίσως η μόνη αιτία ανάκαμψης της έρευνας πάνω στην ICA: στο μεγαλύτερο μέρος της στατιστικής

θεωρίας, θεωρούμε ότι οι τυχαίες μεταβλητές έχουν γκαουσιανές κατανομές, αποκλείοντας έτσι κάθε μέθοδο σχετική με την ICA.

Το Θεώρημα Κεντρικού Ορίου(Central Limit Theorem), ένα κλασικό θεώρημα στη θεωρία πιθανοτήτων, αναφέρει πως το άθροισμα ανεξάρτητων τυχαίων μεταβλητών τείνει προς μια Γκαουσιανή κατανομή, υπό συγκεκριμένες συνθήκες. Έτσι, το άθροισμα δύο ανεξάρτητων τυχαίων μεταβλητών έχει συνήθως μια κατανομή που προσεγγίζει τη Γκαουσιανή περισσότερο από ότι την προσεγγίζουν οι δύο αρχικές ανεξάρτητες μεταβλητές.

Ας υποθέσουμε ότι το διάνυσμα δεδομένων $\mathbf{x}$ είναι ένας συνδυασμός ανεξάρτητων συνιστωσών. Για λόγους απλότητας, υποθέτουμε ότι όλες οι ανεξάρτητες συνιστώσες έχουν πανομοιότυπες κατανομές. Για να υπολογίσουμε μια από τις ανεξάρτητες συνιστώσες, θεωρούμε ένα γραμμικό συνδυασμό των x_i (βλέπε εξίσωση 2.6): ας τον συμβολίσουμε $y = \mathbf{w}^T \mathbf{x} = \sum_i w_i x_i$, όπου $\mathbf{w}$ είναι ένα διάνυσμα που πρέπει να προσδιοριστεί. Αν το $\mathbf{w}$ ήταν μία από τις γραμμές του αντίστροφου του $\mathbf{A}$, τότε αυτός ο γραμμικός συνδυασμός θα ήταν ίσος με τη μία από τις ανεξάρτητες συνιστώσες. Το ερώτημα που προκύπτει είναι πώς θα μπορούσαμε να χρησιμοποιήσουμε το Θεώρημα Κεντρικού Ορίου για να προσδιορίσουμε το $\mathbf{w}$ ώστε να ισούται με μία από τις γραμμές του αντίστροφου του $\mathbf{A}$. Στην πράξη, δεν μπορούμε να προσδιορίσουμε το $\mathbf{w}$ ακριβώς, δεν έχουμε καμία πληροφορία για τον πίνακα $\mathbf{A}$, μπορούμε όμως να βρούμε μία εκτιμήτρια συνάρτηση που δίνει μι καλή προσέγγιση.

Για να δούμε πώς αυτό οδηγεί στη βασική αρχή υπολογισμού της ICA, κάνουμε μία αλλαγή μεταβλητών, ορίζοντας $\mathbf{z} = \mathbf{A}^T \mathbf{w}$. Στη συνέχεια έχουμε $y = \mathbf{w}^T \mathbf{x} = \mathbf{w}^T \mathbf{A} \mathbf{s} = \mathbf{z}^T \mathbf{s}$. Το y είναι συνεπώς ένας γραμμικός συνδυασμός των s_i , με βάρη που δίνονται από τα z_i . Εφόσον ένα άθροισμα ακόμα και δύο ανεξάρτητων τυχαίων μεταβλητών είναι περισσότερο Γκαουσιανό από τις αρχικές μεταβλητές, το $\mathbf{z}^T \mathbf{s}$ είναι περισσότερο Γκαουσιανό από κάθε s_i και γίνεται λιγότερο Γκαουσιανό όταν γίνεται ίσο με το s_i . Σε αυτή την περίπτωση προφανώς, μόνο ένα από τα στοιχεία z_i του $\mathbf{z}$ είναι μη μηδενικό. (Υποθέσαμε ότι τα s_i έχουν πανομοιότυπες κατανομές).

Επομένως, μπορούμε να θεωρήσουμε ως $\mathbf{w}$ ένα διάνυσμα το οποίο να *μεγιστοποιεί την μη προσαρμογή σε Γκαουσιανή κατανομή* του $\mathbf{w}^T \mathbf{x}$. Ένα τέτοιο διάνυσμα θα πρέπει υποχρεωτικά να ανταποκρίνεται(στο σύστημα μετασχηματισμένων μεταβλητών)σε ένα $\mathbf{z}$ το οποίο έχει μόνο μία μη μηδενική συνιστώσα. Αυτό σημαίνει ότι το $\mathbf{w}^T \mathbf{x} = \mathbf{z}^T \mathbf{x}$ ισούται με μία από τις ανεξάρτητες συνιστώσες.

Η μεγιστοποίηση της μη προσαρμογής σε Γκαουσιανή κατανομή του $\mathbf{w}^T \mathbf{x}$ μας δίνει μία από τις ανεξάρτητες συνιστώσες. Στην πραγματικότητα, η

περιοχή βελτιστοποίησης της μη προσαρμογής σε Γκαουσιανή στον n -διάστατο χώρο των $\mathbf{w}$ διανυσμάτων έχει $2n$ τοπικά μέγιστα, δύο για κάθε ανεξάρτητη συνιστώσα, που αντιστοιχούν στα s_i και $-s_i$. Για να βρούμε πολλές ανεξάρτητες μεταβλητές, πρέπει να βρούμε όλα αυτά τα τοπικά μέγιστα. Αυτό δεν είναι δύσκολο, διότι οι διαφορετικές ανεξάρτητες συνιστώσες είναι ασυσχέτιστες: Μπορούμε πάντα να περιορίσουμε την αναζήτηση στο χώρο που δίνει υπολογισμούς ασυσχέτιστους με τους προηγούμενους. Αυτό αντιστοιχεί σε ορθογωνιοποίηση σε ένα κατάλληλα μετασχηματισμένο (π.χ. με προσθήκη λευκού θορύβου) χώρο.

2.4.2 Μεγέθη της μη προσαρμογής σε Γκαουσιανή κατανομή

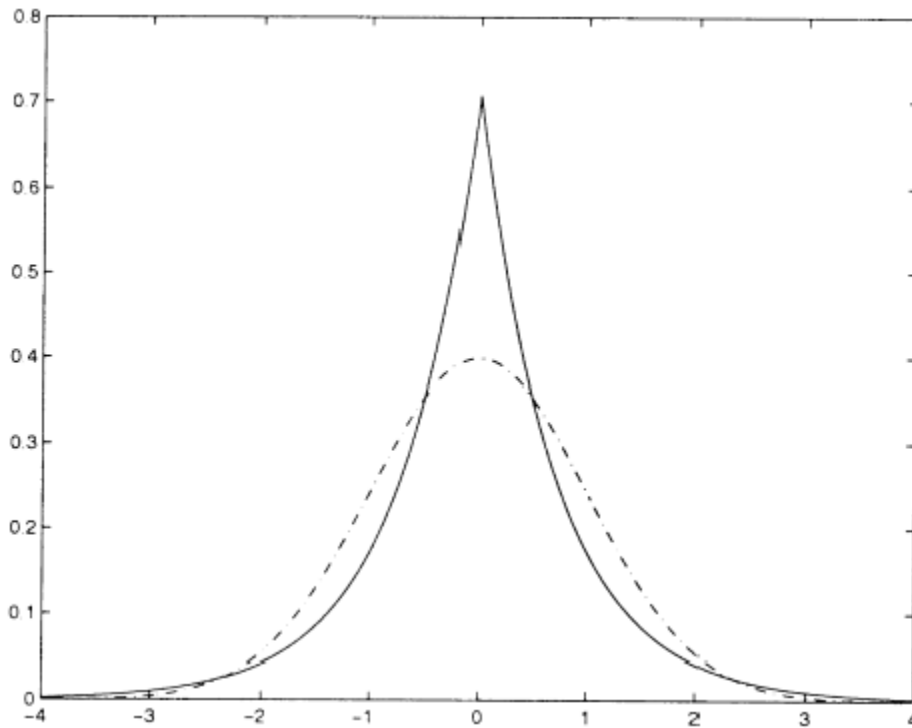
Για να χρησιμοποιήσουμε τη μη προσαρμογή σε Γκαουσιανή κατανομή στον υπολογισμό της ICA, πρέπει να έχουμε ένα ποσοτικό μέγεθος μη προσαρμογής σε Γκαουσιανή κατανομή μιας τυχαίας μεταβλητής, έστω της y . Για να απλοποιήσουμε τα πράγματα, ας υποθέσουμε ότι η y έχει μηδενική μέση τιμή και η διασπορά της ισούται με ένα. Στην πραγματικότητα, μία από τις συναρτήσεις προεπεξεργασίας στους αλγορίθμους της ICA, είναι για να γίνει η απλούστευση αυτή εφικτή.

2.4.2.1 Κύρτωση

Το κλασικό μέγεθος της μη προσαρμογής σε Γκαουσιανή κατανομή είναι η κύρτωση (kurtosis) ή αθροιστής τέταρτης τάξης. Η κύρτωση της y ορίζεται κλασικά ως εξής:

$$kurt(y) = E\{y^4\} - 3(E\{y^2\})^2 \quad (2.16)$$

Αν η y ήταν Γκαουσιανή, η τέταρτη ροπή της θα ισούταν με $3(E\{y^2\})^2$. Συνεπώς, η κύρτωση θα ήταν μηδέν για μία τυχαία Γκαουσιανή μεταβλητή. Στη συγκεκριμένη περίπτωση, αφού έχουμε υποθέσει ότι η y έχει μοναδιαία διασπορά, το δεξιό μέρος απλοποιείται σε $E\{y^4\} - 3$. Αυτό δείχνει ότι η κύρτωση είναι απλά μία κανονικοποιημένη παραλλαγή της τέταρτης ροπής $E\{y^4\}$ και διάφορη του μηδενός. Για τις περισσότερες (όχι όλες) μη Γκαουσιανές τυχαίες μεταβλητές, η κύρτωση είναι μη μηδενική. Η κύρτωση μπορεί να είναι είτε θετική είτε αρνητική. Οι τυχαίες μεταβλητές που έχουν αρνητική κύρτωση ονομάζονται υπο-Γκαουσιανές και αυτές που έχουν θετική κύρτωση ονομάζονται υπερ-Γκαουσιανές. Στη βιβλιογραφία σχετικά με την στατιστική, χρησιμοποιούνται επίσης οι εκφράσεις πλατύκυρτη και λεπτόκυρτη αντίστοιχα. Οι υπερ-Γκαουσιανές τυχαίες μεταβλητές έχουν τυπικά μια «αιχμηρή» σ.π.π. με «βαριές» ουρές. Για παράδειγμα, η σ.π.π. έχει σχετικά υψηλή τιμή στο μηδέν και σε μεγάλες τιμές της μεταβλητής, ενώ έχει μικρές τιμές για ενδιάμεσες τιμές της μεταβλητής.



Εικόνα 2.9 Η συνάρτηση πυκνότητας πιθανότητας (density function) της Laplace κατανομής, που είναι μια τυπική υπερ-Γκαουσιανή κατανομή. Για σύγκριση, η Γκαουσιανή πυκνότητα πιθανότητας δίνεται με διακεκομμένη γραμμή. Και οι δύο πυκνότητες πιθανότητας είναι κανονικοποιημένες σε μοναδιαία διασπορά.

Ένα τυπικό παράδειγμα υπερ-Γκαουσιανής κατανομής είναι η κατανομή Laplace, της οποίας η σ.π.π. (κανονικοποιημένη σε μοναδιαία διασπορά) δίνεται από τη σχέση:

$$p(y) = \frac{1}{\sqrt{2}} \exp(-\sqrt{2} |y|) \quad (2.17)$$

Αυτή η σ.π.π. απεικονίζεται στην Εικόνα 2.9. Οι υπο-Γκαουσιανές τυχαίες μεταβλητές, από την άλλη πλευρά, έχουν τυπικά «επίπεδη» συνάρτηση πυκνότητας πιθανότητας, που είναι συνήθως σταθερή κοντά στο μηδέν και παίρνει πολύ μικρές τιμές για μεγαλύτερες τιμές της μεταβλητής. Ένα τυπικό παράδειγμα είναι η ομοιόμορφη κατανομή στην εξίσωση 2.7.

Τυπικά η μη προσαρμογή σε Γκαουσιανή κατανομή μετρίεται με την απόλυτη τιμή της κύρτωσης. Μπορεί επίσης να χρησιμοποιηθεί το τετράγωνο της κύρτωσης. Αυτό είναι μηδέν για μια Γκαουσιανή μεταβλητή και μεγαλύτερο του μηδενός για τις περισσότερες μη-Γκαουσιανές τυχαίες μεταβλητές. Υπάρχουν μη-Γκαουσιανές τυχαίες μεταβλητές που έχουν μηδενική κύρτωση, αλλά μπορούν να θεωρηθούν πολύ σπάνιες.

Η κύρτωση, ή μάλλον πιο σπάνια η απόλυτη τιμή της, έχει χρησιμοποιηθεί ευρέως σαν ένα μέτρο της μη προσαρμογής σε Γκαουσιανή κατανομή στην ICA και στους σχετικούς τομείς. Ο κυριότερος λόγος είναι η απλότητά της, τόσο η υπολογιστική όσο και η θεωρητική. Η κύρτωση μπορεί να υπολογιστεί απλά χρησιμοποιώντας την τέταρτη ροπή των δεδομένων. Η θεωρητική ανάλυση απλουστεύεται εξαιτίας της ακόλουθης ιδιότητας της γραμμικότητας: αν χ_1 και χ_2 είναι δύο ανεξάρτητες τυχαίες μεταβλητές, τότε ισχύει:

$$kurt(\chi_1 + \chi_2) = kurt(\chi_1) + kurt(\chi_2) \quad (2.18)$$

και

$$kurt(\alpha\chi_1) = \alpha^4 kurt(\chi_1) \quad (2.19)$$

όπου α είναι μία βαθμωτή ποσότητα. Αυτές οι ιδιότητες μπορούν εύκολα να αποδειχθούν χρησιμοποιώντας τον ορισμό.

Για να απεικονίσουμε με ένα απλό παράδειγμα την περιοχή βελτιστοποίησης της κύρτωσης και πώς οι ανεξάρτητες συνιστώσες μπορούν να βρεθούν με την μεγιστοποίηση ή ελαχιστοποίηση της κύρτωσης, ας παρατηρήσουμε το 2-διάστατο μοντέλο $\mathbf{x} = \mathbf{A}\mathbf{s}$.

Κάνοντας ξανά το μετασχηματισμό $\mathbf{z} = \mathbf{A}^T \mathbf{w}$, έχουμε $y = \mathbf{w}^T \mathbf{x} = \mathbf{w}^T \mathbf{A}\mathbf{s} = \mathbf{z}^T \mathbf{s} = z_1 s_1 + z_2 s_2$. Τώρα, με βάση την προσθετική ιδιότητα της κύρτωσης, έχουμε $kurt(y) = kurt(z_1 \chi_1) + kurt(z_2 \chi_2) = z_1^4 kurt(s_1) + z_2^4 kurt(s_2)$. Από την άλλη πλευρά, κάναμε την παραδοχή ότι η διασπορά του y είναι ίση με ένα, βασιζόμενοι στην ίδια υπόθεση αναφορικά με τις s_1, s_2 . Αυτό υποδηλώνει έναν περιορισμό για το $\mathbf{z}$: $E\{y^2\} = z_1^2 + z_2^2 = 1$. Γεωμετρικά, αυτό σημαίνει ότι το διάνυσμα $\mathbf{z}$ περιορίζεται στον μοναδιαίο κύκλο στο 2-δισδιάστατο επίπεδο. Το πρόβλημα βελτιστοποίησης γίνεται τώρα πρόβλημα εύρεσης του μεγίστου της συνάρτησης $|kurt(y)| = |z_1^4 kurt(s_1) + z_2^4 kurt(s_2)|$ στον μοναδιαίο κύκλο. Για λόγους απλότητας, μπορούμε να θεωρήσουμε ότι η κύρτωση έχει ίδιο πρόσημο, και στην περίπτωση αυτή το σύμβολο της απόλυτης τιμής μπορεί να παραληφθεί. Η γραφική παράσταση αυτής της συνάρτησης είναι η «περιοχή βελτιστοποίησης» για το πρόβλημα.

Δεν είναι δύσκολο να αποδειχθεί [16] ότι το μέγιστο είναι στα σημεία όπου ακριβώς ένα από τα στοιχεία του διανύσματος $\mathbf{z}$ είναι μηδέν και τα υπόλοιπα μη μηδενικά. Λόγω του περιορισμού του μοναδιαίου κύκλου, το μη μηδενικό στοιχείο θα πρέπει να είναι ή 1 ή -1. Αλλά αυτά τα σημεία είναι ακριβώς αυτά που προκύπτουν όταν το y ισούται με μία από τις ανεξάρτητες συνιστώσες $\pm s_i$, και το πρόβλημα έτσι έχει λυθεί.

Στην πράξη θα μπορούσαμε να ξεκινήσουμε με κάποιο διάνυσμα βαρών $\mathbf{w}$, να υπολογίσουμε την κατεύθυνση στην οποία η κύρτωση της $y = \mathbf{w}^T \mathbf{x}$ αυξάνεται περισσότερο (αν η κύρτωση είναι θετική) ή μειώνεται περισσότερο (αν η κύρτωση είναι αρνητική) βασιζόμενοι στο διαθέσιμο δείγμα $\mathbf{x}(1), \dots, \mathbf{x}(T)$ του συνδυαστικού διανύσματος $\mathbf{x}$, και να χρησιμοποιήσουμε μία μέθοδο κλίσης (gradient method) ή μία από τις επεκτάσεις της για να βρούμε ένα νέο διάνυσμα $\mathbf{w}$. Το παράδειγμα μπορεί να γενικευθεί σε αυθαίρετες διαστάσεις, δείχνοντας ότι η κύρτωση μπορεί θεωρητικά να χρησιμοποιηθεί ως κριτήριο βελτίωσης για το πρόβλημα της ICA.

Ωστόσο, η κύρτωση έχει και μερικά μειονεκτήματα στην πράξη, όταν η τιμή της πρέπει να υπολογιστεί με βάση ένα ήδη υπολογισμένο δείγμα. Το κύριο πρόβλημα είναι ότι η κύρτωση είναι ευαίσθητη σε παρεκτρεπόμενες τιμές [17]. Η τιμή της μπορεί να εξαρτάται μόνο από λίγες παρατηρήσεις στις ουρές των κατανομών, οι οποίες μπορεί να είναι λανθασμένες ή άσχετες παρατηρήσεις. Με άλλα λόγια, η κύρτωση δεν είναι ένα ισχυρό μέτρο της μη προσαρμογής σε Γκαουσιανή κατανομή.

Συνεπώς, άλλα μέτρα της μη προσαρμογής σε Γκαουσιανή ίσως είναι καλύτερα από την κύρτωση σε ορισμένες περιπτώσεις. Παρακάτω θα αναφερθούμε στην αρνητική εντροπία της οποίας οι ιδιότητες είναι μάλλον αντίθετες από αυτές της κύρτωσης, και τελικά εισάγουν προσεγγίσεις αρνητικής εντροπίας οι οποίες συνδυάζουν τις καλές ιδιότητες και των δύο μεγεθών.

2.4.2.2 Αρνητική Εντροπία

Ένα δεύτερο πολύ σημαντικό μέτρο της μη προσαρμογής σε Γκαουσιανή κατανομή είναι η αρνητική εντροπία. Η αρνητική εντροπία βασίζεται στην ποσότητα της διαφορικής εντροπίας που προέρχεται από τη θεωρία πληροφορίας.

Η εντροπία είναι η βασική έννοια της θεωρίας πληροφοριών. Η εντροπία μιας τυχαίας μεταβλητής μπορεί να ερμηνευθεί ως ο βαθμός πληροφορίας που δίνει η παρατήρηση της μεταβλητής. Όσο πιο «τυχαία», δηλαδή όσο πιο απρόβλεπτη και αδόμητη είναι, τόσο μεγαλύτερη εντροπία έχει. Πιο συγκεκριμένα, η εντροπία σχετίζεται στενά με το μήκος της κωδικοποίησης μιας τυχαίας μεταβλητής και υπό κάποιες υποθέσεις απλούστευσης η εντροπία είναι το μήκος της κωδικοποίησης της τυχαίας μεταβλητής.

Η εντροπία για μία διακριτή τυχαία μεταβλητή Y ορίζεται ως:

$$H(Y) = - \sum_i P(Y = a_i) \log P(Y = a_i) \quad (2.20)$$

όπου τα a_i είναι οι πιθανές τιμές της Y . Αυτός ο πολύ γνωστός ορισμός μπορεί να γενικευθεί για συνεχείς τυχαίες μεταβλητές και διανύσματα, όπου στην περίπτωση αυτή ονομάζεται συχνά διαφορική εντροπία. Η διαφορική εντροπία H ενός τυχαίου διανύσματος $\mathbf{y}$ με πυκνότητα πιθανότητας $f(\mathbf{y})$ ορίζεται ως [18], [19]:

$$H(\mathbf{y}) = - \int f(\mathbf{y}) \log f(\mathbf{y}) d\mathbf{y} \quad (2.21)$$

Ένα θεμελιώδες συμπέρασμα της θεωρίας πληροφοριών είναι ότι μία Γκαουσιανή μεταβλητή έχει τη μεγαλύτερη εντροπία μεταξύ όλων των τυχαίων μεταβλητών ίδιας διασποράς [18], [19]. Αυτό σημαίνει ότι η εντροπία θα μπορούσε να χρησιμοποιηθεί ως μέτρο της μη προσαρμογής σε Γκαουσιανή κατανομή. Στην πραγματικότητα, αυτό δείχνει ότι η Γκαουσιανή κατανομή είναι η «πιο τυχαία» ή η λιγότερο δομημένη από όλες τις κατανομές. Η εντροπία είναι μικρή για κατανομές που είναι καθαρά συγκεντρωμένες σε συγκεκριμένες τιμές, για παράδειγμα όταν η μεταβλητή είναι σαφώς συγκεντρωμένη, ή έχει σ.π.π. που είναι πολύ «αιχμηρή».

Για να αποκτήσουμε ένα μέτρο της μη προσαρμογής σε Γκαουσιανή κατανομή που να είναι μηδέν για μία Γκαουσιανή μεταβλητή και πάντα μη αρνητικό, συχνά χρησιμοποιείται μια ελαφρώς τροποποιημένη παραλλαγή του ορισμού της διαφορικής εντροπίας, που ονομάζεται αρνητική εντροπία. Η αρνητική εντροπία J ορίζεται ως ακολούθως:

$$J(\mathbf{y}) = H(\mathbf{y}_{gauss}) - H(\mathbf{y}) \quad (2.22)$$

όπου $\mathbf{y}_{gauss}$ είναι μία Γκαουσιανή τυχαία μεταβλητή με ίδιο πίνακα συνδιασποράς με την $\mathbf{y}$. Εξαιτίας των ιδιοτήτων που αναφέρθηκαν παραπάνω, η αρνητική εντροπία είναι πάντα μη αρνητική, και είναι μηδέν αν και μόνο αν η $\mathbf{y}$ έχει Γκαουσιανή κατανομή. Η αρνητική εντροπία έχει επιπλέον την ενδιαφέρουσα ιδιότητα ότι είναι αμετάβλητη για αντιστρέψιμους γραμμικούς συνδυασμούς [13], [20].

Το πλεονέκτημα της χρήσης της αρνητικής εντροπίας, ή, ισότιμα, της διαφορικής εντροπίας, ως μέτρο της μη προσαρμογής σε Γκαουσιανή κατανομή, είναι ότι η αρνητική εντροπία είναι καλά τεκμηριωμένη από τη στατιστική θεωρία. Στην πραγματικότητα, η αρνητική εντροπία είναι κατά κάποιο τρόπο η βέλτιστη εκτιμήτρια συνάρτηση της μη προσαρμογής σε Γκαουσιανή κατανομή, όσον αφορά στις στατιστικές ιδιότητες. Το πρόβλημα στη χρήση της αρνητικής εντροπίας είναι, ωστόσο, ότι είναι υπολογιστικά πολύ δύσκολη. Ο υπολογισμός της αρνητικής εντροπίας χρησιμοποιώντας τον ορισμό θα απαιτούσε τον υπολογισμό της (πιθανώς μη παραμετρικής) σ.π.π.. Συνεπώς, πολύ χρήσιμες είναι οι απλούστερες προσεγγίσεις της αρνητικής εντροπίας.

2.4.2.3 Προσεγγίσεις της αρνητικής εντροπίας

Ο υπολογισμός της αρνητικής εντροπίας είναι δύσκολος, όπως αναφέρθηκε παραπάνω, και συνεπώς αυτή η συνάρτηση αντίθεσης παραμένει περισσότερο θεωρητική. Στην πράξη, πρέπει να χρησιμοποιηθούν κάποιες προσεγγίσεις.

Η κλασσική μέθοδος της προσεγγιστικής αρνητικής εντροπίας χρησιμοποιεί υψηλότερης τάξης ροπές, για παράδειγμα [21]:

$$J(y) \approx \frac{1}{12}E\{y^3\}^2 + \frac{1}{48}kurt(y)^2 \quad (2.23)$$

Η τυχαία μεταβλητή θεωρείται μηδενικής μέσης τιμής και μοναδιαίας διασποράς. Ωστόσο, η εγκυρότητα τέτοιων προσεγγίσεων μπορεί να είναι αρκετά περιορισμένη. Πιο συγκεκριμένα, αυτές οι προσεγγίσεις παρουσιάζουν προβλήματα μη-ανθεκτικότητας λόγω της κύρτωσης.

Για την αποφυγή των προβλημάτων που συναντώνται με τις προηγούμενες προσεγγίσεις της αρνητικής εντροπίας, αναπτύχθηκαν νέες προσεγγίσεις. Οι προσεγγίσεις αυτές βασίστηκαν στην αρχή της μέγιστης εντροπίας. Γενικά λαμβάνουμε την ακόλουθη προσέγγιση:

$$J(y) \approx \sum_{i=1}^p k_i [E\{G_i(y)\} - E\{G_i(v)\}]^2 \quad (2.24)$$

όπου k_i είναι θετικές σταθερές και v είναι μία Γκαουσιανή μεταβλητή με μηδενική μέση τιμή και μοναδιαία διασπορά (π.χ. προτυποποιημένη). Η μεταβλητή y θεωρείται μηδενικής μέσης τιμής και μοναδιαίας διασποράς, και οι συναρτήσεις G_i είναι μη-τετραγωνικές συναρτήσεις. Επισημαίνεται ότι ακόμα και στις περιπτώσεις όπου αυτή η προσέγγιση δεν είναι πολύ ακριβής, η εξίσωση 2.24 μπορεί να χρησιμοποιηθεί για να διατυπωθεί ένα μέτρο της μη προσαρμογής σε Γκαουσιανή κατανομή, που είναι συνεπές υπό την έννοια ότι θα είναι πάντα μη αρνητική, και ίση με μηδέν αν η y έχει Γκαουσιανή κατανομή.

Στην περίπτωση που χρησιμοποιούμε μόνο μία μη-τετραγωνική συνάρτηση G , η προσέγγιση γίνεται:

$$J(y) \propto [E\{G(y)\} - E\{G(v)\}]^2 \quad (2.25)$$

Στην πράξη για κάθε μη-τετραγωνική συνάρτηση G . Αυτό είναι σαφώς μία γενίκευση της προσέγγισης που βασίζεται στη ροπή στην εξίσωση 2.23, αν η y είναι συμμετρική. Πράγματι, θέτοντας $G(y) = y^4$, προκύπτει ακριβώς η εξίσωση 2.23, μία προσέγγιση βασιζόμενη στην κύρτωση.

Αλλά το ζήτημα εδώ είναι ότι με κατάλληλη επιλογή του G , προκύπτουν προσεγγίσεις της αρνητικής εντροπίας πολύ καλύτερες από αυτές που δίνονται με την εξίσωση 2.23. Συγκεκριμένα, επιλέγοντας ένα G που δεν αυξάνεται πολύ γρήγορα, προκύπτουν πιο αποδοτικές εκτιμήτριες συναρτήσεις. Οι παρακάτω επιλογές του G έχουν αποδειχθεί πολύ χρήσιμες:

$$G_1(u) = \frac{1}{a_1} \log \cos a_1 u, \quad G_2(u) = -\exp(-u^2/2) \quad (2.26)$$

όπου $1 \leq a_1 \leq 2$ είναι μία κατάλληλη σταθερά.

Έτσι, εξασφαλίζουμε προσεγγίσεις της αρνητικής εντροπίας που δίνουν μία πολύ καλή συμβιβαστική λύση μεταξύ των ιδιοτήτων των δύο κλασσικών μέτρων της μη προσαρμογής σε Γκαουσιανή κατανομή, της κύρτωσης και της αρνητικής εντροπίας. Είναι εννοιολογικά απλές, γρήγορες στον υπολογισμό τους, ωστόσο έχουν εκπληκτικές στατιστικές ιδιότητες, και ειδικά ισχύ.

2.4.3 Ελαχιστοποίηση της αμοιβαίας πληροφορίας

2.4.3.1 Αμοιβαία Πληροφορία

Μία άλλη προσέγγιση για τον υπολογισμό ICA, εμπνευσμένη από τη θεωρία πληροφοριών, είναι η ελαχιστοποίηση της αμοιβαίας πληροφορίας. Χρησιμοποιώντας την έννοια της διαφορικής εντροπίας, ορίζουμε την αμοιβαία πληροφορία I μεταξύ m (βαθμωτών) τυχαίων μεταβλητών $y_i, i = 1 \dots m$ ως ακολούθως:

$$I(y_1, y_2, \dots, y_m) = \sum_{i=1}^m H(y_i) - H(\mathbf{y}) \quad (2.27)$$

Η αμοιβαία πληροφορία είναι φυσικό μέτρο της εξάρτησης μεταξύ τυχαίων μεταβλητών. Στην πραγματικότητα, είναι ισότιμη με τη γνωστή απόκλιση Kullback-Leibler μεταξύ της από κοινού πυκνότητας πιθανότητας $f(\mathbf{y})$ και του γινομένου των περιθωριακών πυκνοτήτων πιθανότητάς της, ένα πολύ φυσικό μέτρο για την ανεξαρτησία. Είναι πάντα μη αρνητική, και μηδενίζεται αν και μόνο αν οι μεταβλητές είναι στατιστικώς ανεξάρτητες. Κατά συνέπεια, η αμοιβαία πληροφορία λαμβάνει υπόψη ολόκληρη τη δομή εξάρτησης των μεταβλητών, και όχι μόνο τη συνδιασπορά, όπως η PCA και άλλες σχετικές μέθοδοι. Η αμοιβαία πληροφορία μπορεί να ερμηνευθεί χρησιμοποιώντας την ερμηνεία της εντροπίας ως μήκος κωδικοποίησης. Οι όροι $H(y_i)$ δίνουν τα μήκη των κωδίκων για τις y_i όταν αυτές κωδικοποιούνται ξεχωριστά, και ο όρος $H(\mathbf{y})$ δίνει το μήκος του κώδικα όταν το $\mathbf{y}$ κωδικοποιείται ως τυχαίο διάνυσμα, π.χ. όλες οι συνιστώσες κωδικοποιούνται στον ίδιο κώδικα. Η αμοιβαία πληροφορία συνεπώς δείχνει τι μείωση του μήκους του κώδικα επιτυγχάνεται κωδικοποιώντας ολόκληρο το διάνυσμα αντί κάθε συνιστώσα ξεχωριστά. Γενικά, λαμβάνονται καλύτεροι κώδικες όταν κωδικοποιείται ολόκληρο το διάνυσμα. Ωστόσο, αν οι y_i είναι ανεξάρτητες, δεν δίνουν καμία πληροφορία η μία για την άλλη και μπορούν να κωδικοποιηθούν ξεχωριστά χωρίς να αυξηθεί το μήκος του κώδικα.

Μία σημαντική ιδιότητα της αμοιβαίας πληροφορίας [18], [19] είναι ότι για ένα αντιστρέψιμο γραμμικό συνδυασμό $\mathbf{y} = \mathbf{W}\mathbf{x}$ έχουμε:

$$I(y_1, y_2, \dots, y_n) = \sum_i H(y_i) - H(\mathbf{x}) - \log|\det \mathbf{W}| \quad (2.28)$$

Ας εξετάσουμε τι θα συμβεί αν εισάγουμε τον περιορισμό οι y_i να είναι ασυσχέτιστες και μοναδιαίας διασποράς. Αυτό σημαίνει

$$E(\mathbf{y}\mathbf{y}^T) = \mathbf{W}E\{\mathbf{x}\mathbf{x}^T\}\mathbf{W}^T = \mathbf{I},$$

το οποίο υποδηλώνει ότι

$$\begin{aligned} \det \mathbf{I} = 1 &= (\det \mathbf{W} E\{\mathbf{x}\mathbf{x}^T\} \mathbf{W}^T) \\ &= (\det \mathbf{W}) \det E\{\mathbf{x}\mathbf{x}^T\} (\det \mathbf{W}^T) \end{aligned} \quad (2.29)$$

και αυτό υποδηλώνει ότι η $\det \mathbf{W}$ πρέπει να είναι σταθερή. Επιπλέον, για y_i μοναδιαίας διασποράς, η εντροπία και η αρνητική εντροπία διαφέρουν μόνο κατά μία σταθερά και το πρόσημο. Συνεπώς έχουμε:

$$I(y_1, y_2, \dots, y_n) = C - \sum_i J(y_i) \quad (2.30)$$

όπου C είναι μία σταθερά που δεν εξαρτάται από το $\mathbf{W}$. Αυτό δείχνει τη βασική σχέση μεταξύ της αρνητικής εντροπίας και της αμοιβαίας πληροφορίας.

2.4.3.2 Ορίζοντας την ICA με βάση την αμοιβαία πληροφορία

Εφόσον η αμοιβαία πληροφορία είναι το φυσικό μέτρο, που βασίζεται στη θεωρία πληροφοριών για την ανεξαρτησία των τυχαίων μεταβλητών, θα μπορούσαμε να τη χρησιμοποιήσουμε ως κριτήριο για την εύρεση του μετασχηματισμού ICA. Σε αυτή την προσέγγιση, που είναι εναλλακτική της προσέγγισης υπολογισμού του μοντέλου, ορίζουμε την ICA τυχαίου διανύσματος $\mathbf{x}$ ως έναν αντίστροφο μετασχηματισμό, όπως στην εξίσωση 2.6, όπου ο πίνακας $\mathbf{W}$ ορίζεται έτσι ώστε η αμοιβαία πληροφορία των μετασχηματισμένων συνιστωσών s_i να ελαχιστοποιείται.

Είναι τώρα προφανές από την εξίσωση 2.30 ότι η εύρεση του αντίστροφου μετασχηματισμού του $\mathbf{W}$ που ελαχιστοποιεί την αμοιβαία πληροφορία είναι κατά προσέγγιση ισότιμη με την εύρεση κατευθύνσεων στις οποίες η αρνητική εντροπία μεγιστοποιείται. Πιο συγκεκριμένα, η εξίσωση 2.30 δείχνει ότι ο υπολογισμός της ICA με ελαχιστοποίηση της αμοιβαίας πληροφορίας είναι ισοδύναμος με τη μεγιστοποίηση του αθροίσματος των «μη-Γκαουσιανοτήτων» των εκτιμητριών συναρτήσεων, όταν οι υπολογισμοί περιορίζονται να είναι ασυσχέτιστοι. Η δέσμευση της μη συσχέτισης στην πραγματικότητα δεν είναι απαραίτητη, αλλά απλοποιεί σημαντικά τους υπολογισμούς, ώστε μετά να μπορεί να χρησιμοποιηθεί η απλούστερη μορφή στην εξίσωση 2.30 αντί της πιο περίπλοκης μορφής στην εξίσωση 2.28.

Συνεπώς, βλέπουμε ότι η διατύπωση της ICA ως ελαχιστοποίηση της αμοιβαίας πληροφορίας δίνει άλλη μία αυστηρή τεκμηρίωση της λιγότερο αυστηρά παρουσιαζόμενης ιδέας της εύρεσης μέγιστων μη-Γκαουσιανών κατευθύνσεων.

2.4.4 Εκτίμηση της Μέγιστης Πιθανοφάνειας

2.4.4.1 Η Πιθανότητα

Μία πολύ δημοφιλής προσέγγιση της εκτίμησης του μοντέλου ICA είναι η εκτίμηση της μέγιστης πιθανοφάνειας, που συνδέεται στενά με την αρχή μεγιστοποίησης της εντροπίας (infomax principle).

Είναι δυνατό να σχηματιστεί άμεσα η πιθανότητα στο χωρίς θόρυβο μοντέλο ICA [22], και μετά να υπολογιστεί το μοντέλο με τη μέθοδο μέγιστης πιθανοφάνειας. Συμβολίζοντας με $\mathbf{W} = (\mathbf{w}_1, \dots, \mathbf{w}_n)^T$ τον πίνακα $\mathbf{A}^{-1}$, η λογαριθμική πιθανοφάνεια λαμβάνει τη μορφή [22]:

$$L = \sum_{i=1}^T \sum_{j=1}^n \log f_i(\mathbf{w}_j^T \mathbf{x}(t)) + T \log |\det \mathbf{W}| \quad (2.31)$$

όπου f_i είναι οι συναρτήσεις πυκνότητας πιθανότητας (density functions) των s_i (εδώ θεωρούνται γνωστά), και το $\mathbf{x}(t), t = 1, \dots, T$ είναι οι πραγματοποιήσεις του $\mathbf{x}$. Ο όρος $\log |\det \mathbf{W}|$ στην πιθανότητα προέρχεται από τον κλασικό κανόνα για τον (γραμμικό) μετασχηματισμό τυχαίων μεταβλητών και των πυκνοτήτων πιθανότητας [19]. Γενικά, για κάθε τυχαίο διάνυσμα $\mathbf{x}$ με πυκνότητα πιθανότητας $p_{\mathbf{x}}$ και για κάθε πίνακα $\mathbf{W}$, η πυκνότητα πιθανότητας των $\mathbf{y} = \mathbf{W}\mathbf{x}$ δίνεται από το $p_{\mathbf{y}}(\mathbf{W}^{-1}\mathbf{y})|\det \mathbf{W}^{-1}|$.

2.4.4.2 Η αρχή της μεγιστοποίησης της εντροπίας (The infomax principle)

Μία άλλη σχετική συνάρτηση αντίθεσης προήλθε από την οπτική γωνία ενός νευρωνικού δικτύου [23], [24]. Αυτή βασίστηκε στη μεγιστοποίηση της εντροπίας εξόδου (ή της ροής της πληροφορίας) ενός νευρωνικού δικτύου με μη-γραμμικές εξόδους. Ας υποθέσουμε ότι το $\mathbf{x}$ είναι είσοδος σε ένα νευρωνικό δίκτυο του οποίου οι εξοδοί είναι της μορφής $\varphi_i(\mathbf{w}_i^T \mathbf{x})$, όπου οι φ_i είναι μη γραμμικές βαθμωτές συναρτήσεις, και τα $\mathbf{w}_i$ είναι τα διανύσματα βαρών των νευρώνων. Θέλουμε να μεγιστοποιήσουμε την εντροπία των εξόδων:

$$L_2 = H(\varphi_1(\mathbf{w}_1^T \mathbf{x}), \dots, \varphi_n(\mathbf{w}_n^T \mathbf{x})) \quad (2.32)$$

Αν τα φ_i είναι επιλεγμένα κατάλληλα, το πλαίσιο αυτό επιτρέπει επίσης τον υπολογισμό του μοντέλου ICA. Πράγματι, έχει αποδειχθεί [Cardoso (1997)], [25] ότι η αρχή της μεγιστοποίησης της εντροπίας του δικτύου, ή "infomax", είναι ισοδύναμη με την εκτίμηση της μέγιστης πιθανοφάνειας. Αυτή η ισοδυναμία προϋποθέτει ότι οι μη-γραμμικότητες φ_i που χρησιμοποιούνται στο νευρωνικό δίκτυο έχουν επιλεγεί ως οι αθροιστικές συναρτήσεις κατανομής που αντιστοιχούν στις πυκνότητες πιθανότητας φ_i , δηλαδή $\varphi_i'(\cdot) = f_i(\cdot)$.

2.4.4.3 Συσχέτιση με την αμοιβαία πληροφορία

Για να φανεί η σχέση μεταξύ πιθανότητας και αμοιβαίας πληροφορίας, θεωρούμε την αναμενόμενη τιμή της λογαριθμικής πιθανοφάνειας:

$$\frac{1}{T}E\{L\} = \sum_{i=1}^n E\{\log f_i(\mathbf{w}_i^T \mathbf{x})\} + \log|\det \mathbf{W}| \quad (2.33)$$

Στην πραγματικότητα, αν οι f_i ήταν ίσες με τις πραγματικές κατανομές του $\mathbf{w}_i^T \mathbf{x}$, ο πρώτος όρος θα ήταν ίσος με $-\sum_i H(\mathbf{w}_i^T \mathbf{x})$. Έτσι, η πιθανότητα θα ήταν ίση με την αρνητική αμοιβαία πληροφορία όπως έχει δοθεί στην εξίσωση 2.28.

Στην πράξη η συσχέτιση είναι ακόμα μεγαλύτερη. Αυτό συμβαίνει γιατί στην πράξη δεν γνωρίζουμε τις κατανομές των ανεξάρτητων συνιστωσών. Μια λογική προσέγγιση θα ήταν να υπολογιστεί η πυκνότητα πιθανότητας του $\mathbf{w}_i^T \mathbf{x}$ ως μέρος της μεθόδου εκτίμησης της Μέγιστης Πιθανοφάνειας, και να χρησιμοποιηθεί ως μία προσέγγιση της πυκνότητας πιθανότητας των s_i . Σε αυτή την περίπτωση, πιθανότητα και αμοιβαία πληροφορία είναι, για πρακτικούς λόγους, ίσες.

Παρ' όλα αυτά, υπάρχει μία μικρή διαφορά που μπορεί να είναι πολύ σημαντική στην πράξη. Το πρόβλημα με την εκτίμηση της μέγιστης πιθανοφάνειας είναι ότι οι πυκνότητες πιθανότητας f_i πρέπει να υπολογιστούν σωστά. Δεν χρειάζεται να υπολογιστούν με κάποια μεγάλη ακρίβεια στην πράξη, αρκεί να εκτιμηθεί αν είναι υπο-ή υπερ-Γκαουσιανές [26], [27], [28]. Σε πολλές περιπτώσεις, έχουμε αρκετή προγενέστερη γνώση στις ανεξάρτητες συνιστώσες και δεν χρειάζεται να εκτιμήσουμε τη φύση τους από τα δεδομένα. Σε κάθε περίπτωση, αν η πληροφορία για τη φύση των ανεξάρτητων μεταβλητών δεν είναι σωστή, η εκτίμηση της μέγιστης πιθανοφάνειας θα δώσει εντελώς λανθασμένα αποτελέσματα. Ωστόσο, πρέπει να δοθεί προσοχή στην εκτίμηση της μέγιστης πιθανοφάνειας. Αντιθέτως, με τη χρήση λογικών μεγεθών μη προσαρμογής στη Γκαουσιανή κατανομή, αυτό το πρόβλημα συνήθως δεν εμφανίζεται.

2.5 Προεπεξεργασία για την ICA

Στο προηγούμενο κεφάλαιο, συζητήθηκαν οι στατιστικές αρχές που διέπουν τις μεθόδους ICA. Πριν από την εφαρμογή ενός αλγορίθμου ICA στα δεδομένα, είναι συνήθως πολύ χρήσιμο να κάνουμε κάποια προεπεξεργασία.

2.5.1 Κεντροθέτηση

Η πιο βασική και αναγκαία προεπεξεργασία είναι η κεντροθέτηση του $\mathbf{x}$, δηλαδή η αφαίρεση του μέσου διανύσματος του $\mathbf{m} = E\{\mathbf{x}\}$, έτσι ώστε το $\mathbf{x}$

να γίνει μία μεταβλητή μηδενικού μέσου όρου. Αυτό σημαίνει ότι το $\mathbf{s}$ είναι επίσης μηδενικού μέσου όρου, όπως μπορεί να φανεί με τη λήψη αναμενόμενων τιμών και στα δύο μέλη της εξίσωσης 2.4.

Αυτή η προεπεξεργασία γίνεται αποκλειστικά και μόνο για την απλοποίηση των αλγορίθμων ICA: δε σημαίνει ότι ο μέσος όρος δεν μπορεί να υπολογιστεί. Μετά τον υπολογισμό του συνδυαστικού πίνακα $\mathbf{A}$ με κεντροθετημένα δεδομένα, μπορούμε να ολοκληρώσουμε τον υπολογισμό με την προσθήκη του μέσου διάνυσματος $\mathbf{s}$ πίσω στους κεντροθετημένους υπολογισμούς του $\mathbf{s}$. Το μέσο διάνυσμα δίνεται από το $\mathbf{A}^{-1}\mathbf{m}$, όπου $\mathbf{m}$ είναι ο μέσος που είχε αφαιρεθεί στην προεπεξεργασία.

2.5.2 Αποσυσχέτιση

Μία άλλη χρήσιμη στρατηγική προεπεξεργασίας στην ICA είναι να «αποσυσχετίζονται» πρώτα οι παρατηρούμενες μεταβλητές. Αυτό σημαίνει ότι πριν από την εφαρμογή του αλγορίθμου ICA (και μετά την κεντροθέτηση), μετασχηματίζουμε το παρατηρούμενο διάνυσμα $\mathbf{x}$ γραμμικά, έτσι ώστε να αποκτήσουμε ένα νέο διάνυσμα $\tilde{\mathbf{x}}$ το οποίο είναι αποσυσχετισμένο, δηλαδή οι συνιστώσες του είναι ασυσχέτιστες και οι διασπορές τους ισούνται με τη μονάδα. Με άλλα λόγια, ο πίνακας συνδιασποράς του $\tilde{\mathbf{x}}$ ισούται με το μοναδιαίο πίνακα:

$$E\{\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T\} = \mathbf{I} \quad (2.34)$$

Ο μετασχηματισμός αποσυσχέτισης είναι πάντα εφικτός. Μία δημοφιλής μέθοδος αποσυσχέτισης είναι η χρήση της αποσύνθεσης ιδιοτιμής (eigenvalue decomposition-EVD) του πίνακα συνδιασποράς $E\{\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T\} = \mathbf{E}\mathbf{D}\mathbf{E}^T$, όπου $\mathbf{E}$ είναι ο ορθογώνιος πίνακας ιδιοδιανυσμάτων του $E\{\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T\}$ και $\mathbf{D}$ είναι ο διαγώνιος πίνακας των ιδιοτιμών του, $\mathbf{D} = \text{diag}(d_1, \dots, d_n)$. Σημειώνουμε ότι το $E\{\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T\}$ μπορεί να υπολογιστεί με συνηθισμένο τρόπο από το διαθέσιμο δείγμα $\mathbf{x}(1), \dots, \mathbf{x}(T)$. Η αποσυσχέτιση μπορεί πλέον να γίνει με την εξίσωση:

$$\tilde{\mathbf{x}} = \mathbf{E}\mathbf{D}^{-1/2}\mathbf{E}^T\mathbf{x} \quad (2.35)$$

όπου ο πίνακας $\mathbf{D}^{-1/2}$ υπολογίζεται από μία απλή πράξη ως προς τις συνιστώσες ως $\mathbf{D}^{-1/2} = \text{diag}(d_1^{-1/2}, \dots, d_n^{-1/2})$. Είναι εύκολο τώρα να διαπιστωθεί ότι $E\{\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T\} = \mathbf{I}$.

Η αποσυσχέτιση μετατρέπει τον συνδυαστικό πίνακα σε έναν νέο πίνακα, $\tilde{\mathbf{A}}$. Έχουμε από τις εξισώσεις 2.4 και 2.35:

$$\tilde{\mathbf{x}} = \mathbf{E}\mathbf{D}^{-1/2}\mathbf{E}^T\mathbf{A}\mathbf{s} = \tilde{\mathbf{A}}\mathbf{s} \quad (2.36)$$

Η χρησιμότητα της αποσυσχέτισης έγκειται στο γεγονός ότι ο νέος συνδυαστικός πίνακας $\tilde{\mathbf{A}}$ είναι ορθογώνιος. Αυτό φαίνεται από την εξίσωση:

$$E\{\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T\} = \tilde{\mathbf{A}}E\{\mathbf{s}\mathbf{s}^T\}\tilde{\mathbf{A}}^T = \tilde{\mathbf{A}}\tilde{\mathbf{A}}^T = \mathbf{I} \quad (2.37)$$

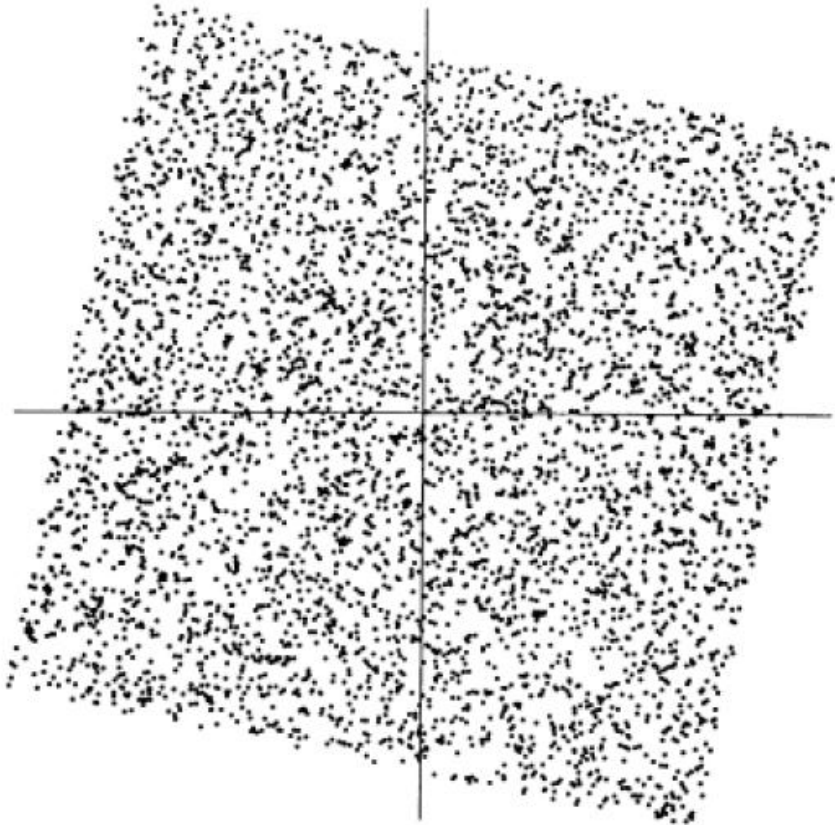
Εδώ βλέπουμε ότι η αποσυσχέτιση μειώνει τον αριθμό των παραμέτρων που πρέπει να υπολογιστούν. Αντί να υπολογιστούν n^2 παράμετροι που είναι τα στοιχεία του αρχικού πίνακα $\mathbf{A}$, χρειάζεται μόνο ο υπολογισμός του νέου ορθογώνιου συνδυαστικού πίνακα $\tilde{\mathbf{A}}$.

Ένας ορθογώνιος πίνακας έχει $n(n-1)/2$ βαθμούς ελευθερίας. Για παράδειγμα, στις δύο διαστάσεις, ένας ορθογώνιος μετασχηματισμός καθορίζεται από μία μόνο παράμετρο γωνίας. Σε περισσότερες διαστάσεις, ένας ορθογώνιος πίνακας έχει περίπου το μισό αριθμό παραμέτρων ενός αυθαίρετου πίνακα. Έτσι, μπορούμε να πούμε ότι η αποσυσχέτιση λύνει «το μισό» πρόβλημα της ICA. Επειδή η αποσυσχέτιση είναι μία πολύ απλή και τυπική διαδικασία, πολύ απλούστερη από τους αλγόριθμους ICA, είναι καλή ιδέα να μειωθεί η πολυπλοκότητα του προβλήματος με αυτόν τον τρόπο.

Είναι επίσης αρκετά χρήσιμο να μειωθεί η διάσταση των δεδομένων την ίδια στιγμή που γίνεται η αποσυσχέτιση. Για να γίνει αυτό εξετάζονται οι ιδιοτιμές d_j του $E\{\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T\}$ και απορρίπτονται αυτές που είναι πολύ μικρές, όπως γίνεται συχνά στην στατιστική τεχνική της ανάλυσης κύριων συνιστωσών. Αυτό συχνά έχει ως αποτέλεσμα τη μείωση του θορύβου. Επιπλέον, η μείωση εμποδίζει την υπερεκπαίδευση (overlearning), η οποία ορισμένες φορές παρατηρείται στην ICA [29].

Μία γραφική απεικόνιση της επίδρασης της αποσυσχέτισης μπορούμε να δούμε στην Εικόνα 2.10, στην οποία τα δεδομένα της Εικόνας 2.7 έχουν αποσυσχετιστεί. Το τετράγωνο που ορίζει τη διασπορά είναι τώρα καθαρά μία περιστρεμμένη έκδοση του αρχικού τετραγώνου της Εικόνας 2.6. Το μόνο που απομένει είναι ο υπολογισμός της γωνίας περιστροφής.

Στο υπόλοιπο του κεφαλαίου αυτού, υποθέτουμε ότι τα δεδομένα έχουν προεπεξεργαστεί με κεντροθέτηση και αποσυσχέτιση. Για απλότητα της σημειογραφίας, συμβολίζουμε τα προεπεξεργασμένα δεδομένα με $\mathbf{x}$, και το μετασχηματισμένο συνδυαστικό πίνακα με $\mathbf{A}$, παραλείποντας την περισπωμένη ($\sim$).



Εικόνα 2.10 Η από κοινού κατανομή των αποσυσχετισμένων (whitened) συνδυασμών.

2.5.3 Περαιτέρω προεπεξεργασία

Η επιτυχία της ICA για ένα συγκεκριμένο σύνολο δεδομένων μπορεί να βασίζεται σημαντικά στην εκτέλεση κάποιων σταδίων προεπεξεργασίας εξαρτώμενων από την εκάστοτε εφαρμογή. Για παράδειγμα, αν τα δεδομένα αποτελούνται από σήματα χρόνου, κάποιο ζωνοπεραστό φιλτράρισμα μπορεί να είναι πολύ χρήσιμο. Σημειώνουμε ότι αν φιλτράρουμε γραμμικά τα παρατηρούμενα σήματα $x_i(t)$ για τη λήψη νέων σημάτων, έστω $x_i^*(t)$ το μοντέλο ICA ακόμη ισχύει για τα $x_i^*(t)$, με τον ίδιο πίνακα μίξης.

Αυτό μπορεί να θεωρηθεί ως εξής: συμβολίζουμε με $\mathbf{X}$ τον πίνακα που περιέχει τις παρατηρήσεις $\mathbf{x}(1)$, ..., $\mathbf{x}(T)$ ως στήλες, και όμοια για τον $\mathbf{S}$. Στη συνέχεια το μοντέλο ICA μπορεί να εκφραστεί ως:

$$\mathbf{X} = \mathbf{AS} \quad (2.38)$$

Τώρα, χρονικό φιλτράρισμα του $\mathbf{X}$ αντιστοιχεί σε πολλαπλασιασμό του $\mathbf{X}$ από δεξιά με έναν πίνακα, ας τον ονομάσουμε $\mathbf{M}$. Αυτό δίνει:

$$\mathbf{X} = \mathbf{XM} = \mathbf{ASM} = \mathbf{AS}^* \quad (2.39)$$

Το οποίο δείχνει ότι το μοντέλο ICA συνεχίζει να ισχύει.

2.6 Ο αλγόριθμος FastICA

Στα προηγούμενα κεφάλαια, εισαγάγαμε διαφορετικά μέτρα της μη-Γκαουσιανής κατανομής, π.χ. αντικειμενικές συναρτήσεις για τον υπολογισμό της ICA. Στην πράξη, χρειαζόμαστε έναν αλγόριθμο για τη μεγιστοποίηση της συνάρτησης αντίθεσης, για παράδειγμα αυτός της εξίσωσης 2.25. Σε αυτό το κεφάλαιο, εισάγουμε μία πολύ επαρκή μέθοδο μεγιστοποίησης. Υποθέτουμε ότι τα δεδομένα έχουν προεπεξεργαστεί με κεντροθέτηση και αποσυσχέτιση, όπως αναφέρθηκε στο προηγούμενο κεφάλαιο.

2.6.1 Περιγραφή λειτουργίας του FastICA σε μία μονάδα νευρωνικών δικτύων

Αρχικά, θα παρουσιάσουμε την εκδοχή μιας μονάδας FastICA. Με τον όρο «μονάδα» αναφερόμαστε σε μία υπολογιστική μονάδα, έναν τεχνητό νευρώνα, που έχει ένα διάνυσμα βάρους $\mathbf{w}$ το οποίο ο νευρώνας αυτός έχει τη δυνατότητα να το ενημερώσει με έναν κανόνα εκμάθησης. Ο κανόνας εκμάθησης FastICA βρίσκει μία κατεύθυνση, δηλαδή μία διανυσματική μονάδα $\mathbf{w}$, έτσι ώστε η μη-Γκαουσιανή προβολή $\mathbf{w}^T \mathbf{x}$ να μεγιστοποιείται. Η μη-Γκαουσιανή μετρίεται κατά προσέγγιση της αρνητικής εντροπίας $J(\mathbf{w}^T \mathbf{x})$ που δίνεται από την εξίσωση 2.25. Η διαφορά $\mathbf{w}^T \mathbf{x}$, πρέπει να περιοριστεί στη μονάδα· για τα αποσυσχετισμένα δεδομένα αυτό είναι ισοδύναμο με τον περιορισμό του κανόνα του $\mathbf{w}$ να είναι μονάδα.

Η μέθοδος FAST-ICA είναι βασισμένη σε ένα σχέδιο επανάληψης σταθερών σημείων (fixed-point), για την εύρεση ενός μέγιστου της μη-Γκαουσιανής κατανομής του $\mathbf{w}^T \mathbf{x}$, όπως υπολογίζεται στην εξίσωση 4.10 [30], [27]. Μπορεί επίσης να θεωρηθεί ως μία προσεγγιστική επανάληψη Newton [30]. Από το g φαίνεται η παράγωγος της μη τετραγωνικής (non-quadratic) συνάρτησης G που χρησιμοποιείται στην εξίσωση 2.25· για παράδειγμα οι παράγωγοι των συναρτήσεων της εξίσωσης 2.26 είναι:

$$g_1(u) = \tanh(a_1 u), \quad g_2 = u \exp\left(-\frac{u^2}{2}\right) \quad (2.40)$$

όπου $1 \leq a_1 \leq 2$ είναι κάποια κατάλληλη σταθερά, που λαμβάνεται συχνά ως $a_1 = 1$. Η βασική μορφή του αλγορίθμου FastICA είναι:

1. Επιλέγουμε ένα αρχικό (π.χ. τυχαίο) διανυσματικό βάρος $\mathbf{w}$.
2. Έστω $\mathbf{w}^+ = E\{\mathbf{x}g(\mathbf{w}^T \mathbf{x})\} - E\{g(\mathbf{w}^T \mathbf{x})\}\mathbf{w}$

3. Έστω $\mathbf{w} = \mathbf{w}^+ / \|\mathbf{w}^+\|$

4. Αν δεν συγκλίνει, επιστρέφουμε στο 2.

Σημειώνουμε ότι η σύγκλιση σημαίνει ότι οι παλιές και νέες τιμές του σημείου $\mathbf{w}$ είναι στην ίδια κατεύθυνση, δηλαδή το σημείο των γινομένων τους είναι (σχεδόν) ίσο με 1. Δεν είναι απαραίτητο το διάνυσμα συγκλίνει σε ένα ενιαίο σημείο, δεδομένου ότι το $\mathbf{w}$ και το $-\mathbf{w}$ καθορίζουν ίδια κατεύθυνση. Αυτό γίνεται επειδή οι ανεξάρτητες συνιστώσες μπορούν να καθοριστούν μόνο μέχρι ένα πολλαπλασιασμένο σήμα. Σημειώνουμε επίσης ότι τα δεδομένα υποτίθεται ότι είναι προ-αποσυσχετισμένα.

Η παραγωγή του FAST-ICA είναι η ακόλουθη. Καταρχήν τα μέγιστα της προσέγγισης της αρνητικής εντροπίας του $\mathbf{w}^T \mathbf{x}$ βρίσκονται σε ορισμένα βέλτιστα του $E\{G(\mathbf{w}^T \mathbf{x})\}$. Σύμφωνα με τους όρους Kuhn-Tucker [31], τα βέλτιστα από το $E\{G(\mathbf{w}^T \mathbf{x})\}$, υπό τον περιορισμό $E\{(\mathbf{w}^T \mathbf{x})^2\} = \|\mathbf{w}\|^2 = 1$ έχουν αποκτηθεί στα σημεία όπου:

$$E\{\mathbf{x}g(\mathbf{w}^T \mathbf{x})\} - \beta \mathbf{w} = 0 \quad (2.41)$$

Ας προσπαθήσουμε να λύσουμε αυτήν την εξίσωση με τη μέθοδο Newton. Δείχνουμε τη συνάρτηση στην αριστερή πλευρά της εξίσωσης 2.41 από το F και λαμβάνουμε τον Jacobian πίνακα του $JF(\mathbf{w})$ ως:

$$JF(\mathbf{w}) = E\{\mathbf{x}\mathbf{x}^T \dot{g}(\mathbf{w}^T \mathbf{x})\} - \beta \mathbf{I} \quad (2.42)$$

Για να απλοποιήσουμε τον αντίστροφο αυτού του πίνακα, προσεγγίζουμε τον πρώτο όρο της εξίσωσης 2.42. Δεδομένου ότι τα δεδομένα είναι σφαιρικά, μια λογική προσέγγιση φαίνεται να είναι η $E\{\mathbf{x}\mathbf{x}^T \dot{g}(\mathbf{w}^T \mathbf{x})\} \approx E\{\mathbf{x}\mathbf{x}^T\}E\{\dot{g}(\mathbf{w}^T \mathbf{x})\} = E\{\dot{g}(\mathbf{w}^T \mathbf{x})\}\mathbf{I}$.

Κατά συνέπεια ο Jacobian πίνακας γίνεται διαγώνιος και μπορεί εύκολα να είναι αντιστρέψιμος. Έτσι λαμβάνουμε την ακόλουθη προσεγγιστική επανάληψη Newton:

$$\mathbf{w}^+ = \mathbf{w} - [E\{\mathbf{x}g(\mathbf{w}^T \mathbf{x})\} - \frac{\beta \mathbf{w}}{[E\{\dot{g}(\mathbf{w}^T \mathbf{x})\} - \beta]}] \quad (2.43)$$

Αυτός ο αλγόριθμος μπορεί να απλοποιηθεί περαιτέρω με τον πολλαπλασιασμό και των δύο πλευρών της εξίσωσης 2.43 από το $\beta - E\{\dot{g}(\mathbf{w}^T \mathbf{x})\}$. Αυτό δίνει, κατόπιν αλγεβρικής απλοποίησης, την επανάληψη FAST-ICA.

Στην πράξη, οι στατιστικοί υπολογισμοί με FAST-ICA πρέπει να αντικατασταθούν από τους υπολογισμούς τους. Οι φυσικοί υπολογισμοί είναι φυσικά τα αντίστοιχα μέσα των δειγμάτων. Ιδανικά, όλα τα διαθέσιμα δεδομένα πρέπει να χρησιμοποιούνται, αλλά αυτό δεν είναι συχνά μια καλή ιδέα επειδή οι υπολογισμοί μπορούν να γίνουν αρκετά απαιτητικοί. Επιπλέον, οι μέσοι όροι μπορούν να υπολογιστούν χρησιμοποιώντας ένα μικρότερο δείγμα, το μέγεθος του οποίου μπορεί να έχει μια ιδιαίτερη επίδραση στην ακρίβεια των τελικών εκτιμήσεων. Τα σημεία δειγματοληψίας πρέπει να επιλεγούν χωριστά σε κάθε

επανάληψη. Αν η σύγκλιση δεν είναι ικανοποιητική, τότε θα μπορούσαμε να αυξήσουμε το μέγεθος του δείγματος.

2.6.2 Περιγραφή λειτουργίας του αλγόριθμου FastICA σε πολλές μονάδες νευρωνικών δικτύων

Στην προηγούμενη ενότητα, ο αλγόριθμος μιας μονάδας υπολογίζει ακριβώς μία από τις ανεξάρτητες συνιστώσες ή μια κατεύθυνση αναζήτησης προβολής. Για να υπολογιστούν οι διάφορες ανεξάρτητες συνιστώσες, χρειάζεται να τρέξουμε τον αλγόριθμο μιας μονάδας FAST-ICA που χρησιμοποιεί αρκετές μονάδες (π.χ. νευρώνες) με διανύσματα βάρους $\mathbf{w}_1, \dots, \mathbf{w}_n$.

Για να αποτρέψουμε τα διαφορετικά διανύσματα από τη σύγκλιση στα ίδια μέγιστα πρέπει να αποσυσχετίσουμε τις εξόδους $\mathbf{w}_1^T \mathbf{x}, \dots, \mathbf{w}_n^T \mathbf{x}$ μετά από κάθε επανάληψη.

Παρουσιάζουμε τρεις μεθόδους για το πώς επιτυγχάνεται αυτό. Σημειώνουμε ότι το αποσυσχετισμένο $\mathbf{x}$, είναι τόσο ασυσχέτιστο όσο και ορθογωνικοποιημένο.

Ένας απλός τρόπος πραγματοποίησης της αποσυσχέτισης είναι ένα σχεδιάγραμμα βασισμένο σε ένα ασυσχέτιστο Gram-Schmidt. Αυτό σημαίνει ότι υπολογίζουμε τις ανεξάρτητες συνιστώσες μία προς μία. Όταν υπολογίσουμε τις ανεξάρτητες συνιστώσες p ή τα διανύσματα p $\mathbf{w}_1, \dots, \mathbf{w}_p$, τρέχουμε το σταθερό σημείο μιας μονάδας του αλγόριθμου για $\mathbf{w}_{p+1}$, και μετά από κάθε βήμα επανάληψης αφαιρούμε από το $\mathbf{w}_{p+1}$ τις «προβολές» $\mathbf{w}_{p+1}^T \mathbf{w}_j \mathbf{w}_j, j = 1, \dots, p$ από τους προηγούμενους υπολογισμούς των διανυσμάτων p και στη συνέχεια απορρίπτουμε το $\mathbf{w}_{p+1}$:

$$1. \text{ Έστω } \mathbf{w}_{p+1} = \mathbf{w}_{p+1} - \sum_{j=1}^p \mathbf{w}_{p+1}^T \mathbf{w}_j \mathbf{w}_j \quad (2.44)$$

$$2. \text{ Έστω } \mathbf{w}_{p+1} = \mathbf{w}_{p+1} / \sqrt{\mathbf{w}_{p+1}^T \mathbf{w}_{p+1}}$$

Σε ορισμένες εφαρμογές, ωστόσο, μπορεί να χρησιμοποιηθεί η συμμετρική αποσυσχέτιση, στην οποία κανένα διάνυσμα δεν είναι «υπερτερεί των άλλων» [32]. Αυτό μπορεί να επιτευχθεί, π.χ. με την κλασική μέθοδο που περιλαμβάνει τις τετραγωνικές ρίζες πινάκων,

$$\text{Έστω } \mathbf{W} = (\mathbf{W}\mathbf{W}^T)^{-\frac{1}{2}} \mathbf{W} \quad (2.45)$$

Όπου $\mathbf{W}$ είναι ο πίνακας $(\mathbf{w}_1, \dots, \mathbf{w}_n)^T$ των διανυσμάτων και η αντίστροφη τετραγωνική ρίζα $(\mathbf{W}\mathbf{W}^T)^{-1/2}$ λαμβάνεται από τις ιδιοτιμές ανάλυσης του $\mathbf{W}\mathbf{W}^T = \mathbf{F} \mathbf{\Lambda} \mathbf{F}^T$ ως $(\mathbf{W}\mathbf{W}^T)^{-\frac{1}{2}} = \mathbf{F} \mathbf{\Lambda}^{-1/2} \mathbf{F}^T$. Μία απλούστερη εναλλακτική λύση είναι ο ακόλουθος επαναληπτικός αλγόριθμος [30],

$$1. \text{ Έστω } \mathbf{W} = \frac{\mathbf{w}}{\sqrt{\|\mathbf{w}\mathbf{w}^T\|}} \quad (2.46)$$

Επαναλαμβάνουμε το βήμα 2 μέχρι τη σύγκλιση:

$$2. \text{ Έστω } \mathbf{W} = \frac{3}{2}\mathbf{W} - \frac{1}{2}\mathbf{W}\mathbf{W}^T\mathbf{W}$$

Ο κανόνας στο βήμα 1 μπορεί να είναι σχεδόν οποιοσδήποτε συνηθισμένος κανόνας πινάκων, π.χ. ο κανόνας 2 ή η μεγαλύτερη απόλυτη τιμή σειρών ή στηλών (αλλά όχι ο κανόνας Frobenius).

2.6.3 Σχέση FastICA και Μέγιστης Πιθανότητας

Τέλος, δίνουμε μια εκδοχή της μεθόδου FastICA που παρουσιάζει σαφώς τη σύνδεση της μέγιστης πληροφορίας (infomax) ή τη μέγιστη πιθανότητα του αλγόριθμου που εισάγεται από τους Amari, Cichocki, Yang, Bell, Sejnowski, Cardoso, Laheld, Cichocki και Unbehauen. Αν εκφράσουμε τον FastICA χρησιμοποιώντας την εξίσωση 2.43 και τον εκφράσουμε με μορφή πινάκων, βλέπουμε ότι λαμβάνει την ακόλουθη μορφή:

$$\mathbf{W}^+ = \mathbf{W} + \text{diag}(a_i)[\text{diag}(\beta_i) + E\{g(\mathbf{y})\mathbf{y}^T\}]\mathbf{W} \quad (2.47)$$

όπου $\mathbf{y}=\mathbf{W}\mathbf{x}$, $\beta_i = -E\{y_i g(y_i)\}$, και $a_i = -1/(\beta_i + E\{\dot{g}(y_i)\})$. Ο πίνακας $\mathbf{W}$ πρέπει να είναι ορθογώνιος μετά από κάθε βήμα. Σε αυτήν την εκδοχή πινάκων, είναι φυσικά ορθογώνια συμμετρικός.

Η παραπάνω εκδοχή του FastICA θα μπορούσε να συγκριθεί με τη στοχαστική μέθοδο κλίσης (stochastic gradient method) για τη μέγιστη πιθανότητα [23], [26], [33], [34]:

$$\mathbf{W}^+ = \mathbf{W} + \mu[\mathbf{I} + g(\mathbf{y})\mathbf{y}^T]\mathbf{W} \quad (2.48)$$

όπου μ είναι το ποσοστό εκμάθησης, όχι απαραίτητα η σταθερά χρόνου. Το g είναι μια συνάρτηση της pdf των ανεξάρτητων συνιστωσών: $g=f'_i/f_i$, όπου f_i είναι η pdf μιας ανεξάρτητης συνιστώσας.

Συγκρίνοντας τις εξισώσεις 2.47 και 2.48, βλέπουμε ότι ο FastICA μπορεί να θεωρηθεί ως αλγόριθμος σταθερών σημείων για τη μέγιστη πιθανότητα εκτίμησης του προτύπου δεδομένων ICA [35]. Στον FastICA, η ταχύτητα σύγκλισης βελτιστοποιείται από την επιλογή των πινάκων $\text{diag}(a_i)$ και $\text{diag}(\beta_i)$. Ένα άλλο πλεονέκτημα του FastICA είναι ότι μπορεί να υπολογίσει τις υπό-Γκαουσιανές και υπέρ-Γκαουσιανές ανεξάρτητες συνιστώσες, κάτι που έρχεται σε αντίθεση με τους συνηθισμένους αλγόριθμους ML, οι οποίοι λειτουργούν μόνο για μια δεδομένη κατηγορία κατανομών.

2.6.4 Ιδιότητες του αλγόριθμου FastICA

Ο αλγόριθμος FAST-ICA έχει κάποιες επιθυμητές ιδιότητες όταν συγκρίνεται με τις υπάρχουσες μεθόδους ICA.

1. Η σύγκλιση είναι κυβική (ή τουλάχιστον τετραγωνική), υπό την προϋπόθεση του προτύπου δεδομένων ICA. Αυτό έρχεται σε αντίθεση με τους συνηθισμένους αλγόριθμους ICA βασισμένους στις (στοχαστικές) μεθόδους καθόδου κλίσης (gradient descent methods), όπου η σύγκλιση είναι μόνο γραμμική. Αυτό σημαίνει μία πολύ γρήγορη σύγκλιση, όπως έχει επιβεβαιωθεί από τις προσομοιώσεις και τα πειράματα σε πραγματικά δεδομένα.
2. Αντίθετα με τους αλγορίθμους που βασίζονται στην κλίση, δεν υπάρχουν παράμετροι βήματος για να επιλεγθούν. Αυτό σημαίνει ότι ο αλγόριθμος είναι εύχρηστος.
3. Ο αλγόριθμος βρίσκει κατευθείαν τις ανεξάρτητες συνιστώσες (πρακτικά) οποιασδήποτε μη-Γκαουσιανής κατανομής χρησιμοποιώντας οποιοδήποτε μη γραμμικό (non-linearity) g . Αυτό έρχεται σε αντίθεση με πολλούς αλγορίθμους, όπου πρέπει πρώτα να γίνει κάποιος υπολογισμός της συνάρτησης κατανομής πιθανότητας και η μη γραμμικότητα πρέπει να επιλεγθεί αναλόγως.
4. Η απόδοση της μεθόδου μπορεί να βελτιστοποιηθεί επιλέγοντας μια κατάλληλη μη γραμμικότητα g . Ειδικότερα, μπορεί να ληφθούν οι αλγόριθμοι που είναι εύρωστοι ή ελάχιστης μεταβλητότητας. Στην πραγματικότητα, οι δύο μη γραμμικότητες της εξίσωσης 2.40 έχουν μερικές βέλτιστες ιδιότητες.
5. Οι ανεξάρτητες συνιστώσες μπορούν να υπολογιστούν μία προς μία, οι οποίες είναι κατά προσέγγιση ισοδύναμες με τη διεξαγωγή αναζήτησης προβολής. Αυτό είναι χρήσιμο στη διερευνητική ανάλυση δεδομένων και μειώνεται ο υπολογιστικός φόρτος της μεθόδου σε περιπτώσεις όπου μόνο μερικές από τις ανεξάρτητες συνιστώσες πρέπει να υπολογιστούν.
6. Ο FastICA έχει τα περισσότερα πλεονεκτήματα των νευρωνικών αλγορίθμων, δηλαδή: είναι παράλληλος, κατανεμημένος, υπολογιστικά απλός και απαιτεί αρκετά μικρό χώρο μνήμης. Οι στοχαστικές μέθοδοι κλίσης φαίνεται να προτιμούνται μόνο αν απαιτείται η γρήγορη προσαρμογή σε ένα μεταβαλλόμενο περιβάλλον.

2.7 Εφαρμογές της ICA

2.7.1 Διαχωρισμός των παράσιτων (artifacts) στα δεδομένα MEG

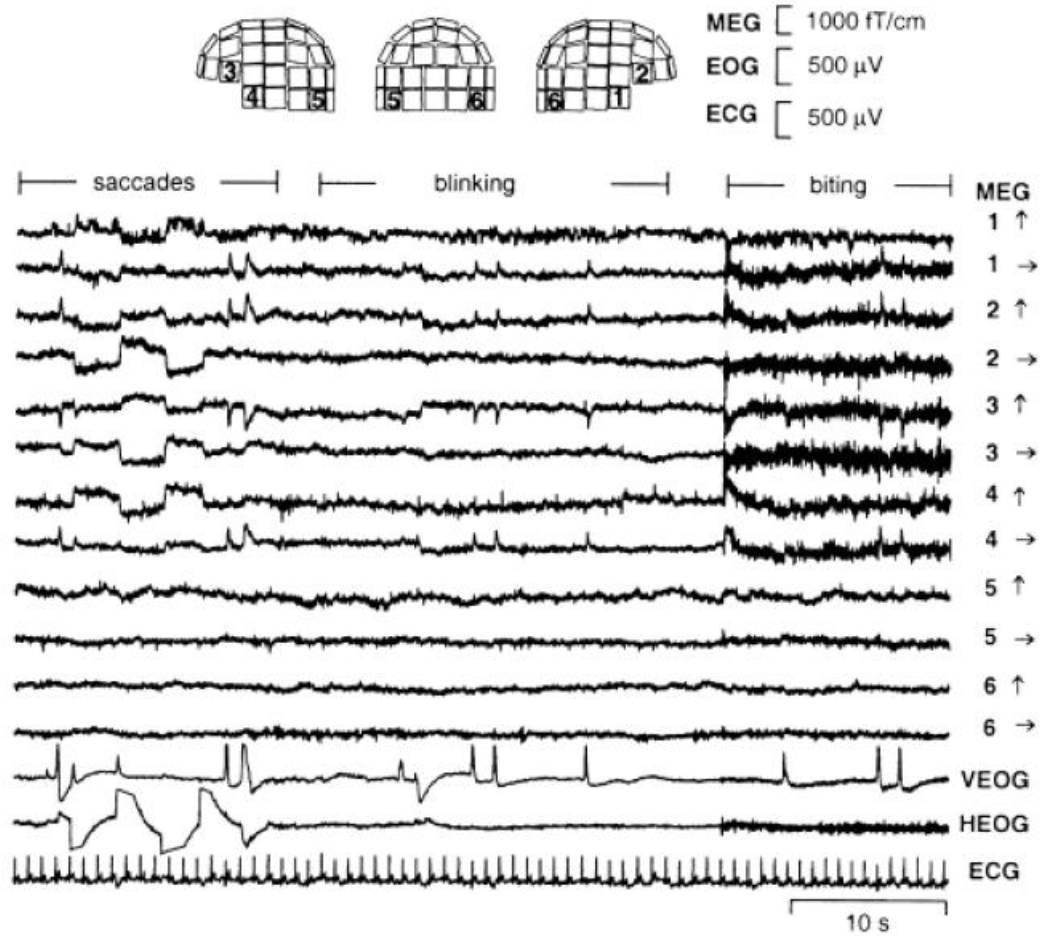
Η Μαγνητοεγκεφαλογραφία (MEG-Magnetoencephalography) είναι μία μη επεμβατική τεχνική με την οποία η δραστηριότητα ή οι νευρώνες του φλοιού μπορούν να μετρηθούν με πολύ καλή χρονική ανάλυση και μέτρια χωρική ανάλυση. Όταν χρησιμοποιείται μία εγγραφή MEG, ως ερευνητικό ή κλινικό εργαλείο, ο ερευνητής μπορεί να αντιμετωπίσει το πρόβλημα της εξαγωγής των απαραίτητων χαρακτηριστικών των νευρομαγνητικών σημάτων, με την παρουσία παρασίτων. Το πλάτος των διαταραχών μπορεί να είναι

μεγαλύτερο από αυτό των εγκεφαλικών σημάτων και τα παράσιτα μπορεί να μοιάζουν στο σχήμα με παθολογικά σήματα.

Μία μέθοδος [36] για το διαχωρισμό της εγκεφαλικής δραστηριότητας από τα παράσιτα χρησιμοποιώντας την ICA βασίζεται στην υπόθεση ότι η εγκεφαλική δραστηριότητα και τα παράσιτα, π.χ. οι κινήσεις των ματιών ή το ανοιγοκλείσιμο των βλεφάρων, ή δυσλειτουργίες του αισθητήρα, είναι από άποψη ανατομίας και φυσιολογίας ξεχωριστές διαδικασίες. Αυτός ο διαχωρισμός αντικατοπτρίζεται στη στατιστική ανεξαρτησία μεταξύ των μαγνητικών σημάτων που παράγονται από αυτές τις διαδικασίες.

Τα σήματα MEG καταγράφονται σε μαγνητικά θωρακισμένο δωμάτιο με ένα νευρομαγνητόμετρο 122 καναλιών που εφαρμόζεται σε ολόκληρο το εξωτερικό μέρος του κεφαλιού και ονομάζεται Neuromag-122. Αυτή η συσκευή συλλέγει δεδομένα σε 61 θέσεις πάνω στο κρανίο, χρησιμοποιώντας ορθογώνια πηνία λήψης διπλού βρόχου (orthogonal double-loop pick-up coils) που συνδέονται σε μία τοπική πηγή ακριβώς από κάτω. Ο εξεταζόμενος κλήθηκε να ανοιγοκλείσει τα βλέφαρα και να κάνει οριζόντιες σπασμωδικές κινήσεις των ματιών (saccades), με σκοπό την παραγωγή τυπικών (οφθαλμικών) παράσιτων. Επιπλέον, για την παραγωγή μυογραφικών (μυϊκών) παράσιτων, ο εξεταζόμενος κλήθηκε να δαγκώσει τα δόντια του για 20 δευτερόλεπτα. Επιπλέον, ένα ακόμη παράσιτο δημιουργήθηκε από ένα ψηφιακό ρολόι που βρισκόταν 1 μέτρο μακριά από την κάσκα στο θωρακισμένο δωμάτιο.

Η Εικόνα 2.11 παρουσιάζει ένα υποσύνολο 12 πηγαίων σημάτων της MEG $x_i(t)$ από τις μετωπιαίες, κροταφικές και ινιακές περιοχές [37]. Η εικόνα επίσης δείχνει τις θέσεις των αντίστοιχων αισθητήρων στην κάσκα. Λόγω της διάστασης των δεδομένων (καταγράφηκαν 122 μαγνητικά σήματα), είναι πρακτικώς αδύνατο να αναπαρασταθούν γραφικά όλα τα σήματα της MEG $x_i(t)$, $i=1, \dots, 122$. Επίσης υπάρχουν δύο κανάλια ηλεκτροοφθαλμογραφήματος και το ηλεκτροκαρδιογράφημα αλλά δεν χρησιμοποιήθηκαν στον υπολογισμό της ICA.

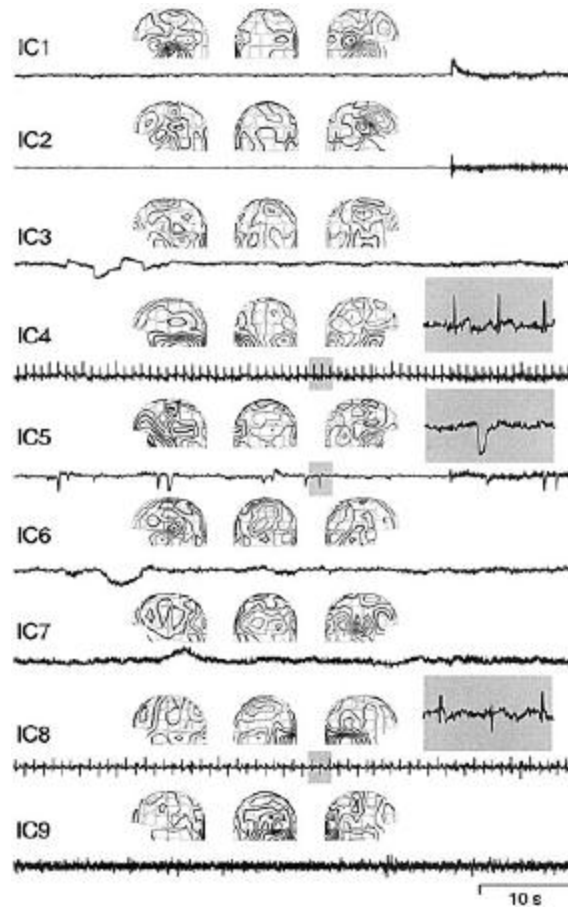


Εικόνα 2.11 Δείγματα σημάτων MEG, που δείχνουν παράσιτα από το ανοιγοκλείσιμο των βλεφάρων (*blinking*), από τις σπασμωδικές κινήσεις των ματιών (*saccades*), από το δάγκωμα (*biting*) και από τον καρδιακό κύκλο. Για κάθε μία από τις έξι θέσεις που φαίνονται στην εικόνα, αποτυπώνονται στο διάγραμμα οι δύο ορθογώνιες κατευθύνσεις των αισθητήρων. (Από [36])

Το διάνυσμα σήματος $\mathbf{x}$ στο μοντέλο ICA (εξίσωση 2.4) αποτελείται τώρα από τα πλάτη $x_i(t)$ των 122 σημάτων σε μία συγκεκριμένη χρονική στιγμή, οπότε η διάσταση είναι $n=122$. Στο θεωρητικό μοντέλο, το $\mathbf{x}$ θεωρείται τυχαίο διάνυσμα, και οι μετρήσεις $\mathbf{x}(t)$ δίνουν ένα σύνολο πραγματοποιήσεων του $\mathbf{x}$ όσο προχωράει ο χρόνος. Σημειώνουμε ότι στο βασικό μοντέλο της ICA που χρησιμοποιούμε, οι χρονικές συσχετίσεις των σημάτων δεν χρησιμοποιούνται καθόλου.

Τα διανύσματα $\mathbf{x}(t)$ αποσυσχετίστηκαν με χρήση της PCA και ταυτόχρονα μειώθηκε η διαστατικότητα. Στη συνέχεια, χρησιμοποιώντας τον αλγόριθμο FastICA, υπολογίστηκε ένα υποσύνολο των γραμμών του πίνακα διαχωρισμού $\mathbf{W}$ της εξίσωσης 2.6. Μόλις ένα διάνυσμα $\mathbf{w}_i$ γίνει διαθέσιμο, ένα

σήμα ICA $s_i(t) = \mathbf{w}_i^T \mathbf{x}(t)$ με το $\mathbf{x}(t)$ να συμβολίζει τώρα το αποσυσχετισμένο και μειωμένων διαστάσεων διάνυσμα σήματος.



Εικόνα 2.12 Εννέα ανεξάρτητες συνιστώσες που βρέθηκαν από δεδομένα MEG. Για κάθε συνιστώσα φαίνονται οι αριστερά, πίσω και δεξιά όψεις των προτύπων δυναμικών που δημιουργούνται από αυτές τις συνιστώσες-η συνεχόμενη γραμμή αντιπροσωπεύει τη μαγνητική ροή που εξέρχεται από το κεφάλι, και η διακεκομμένη γραμμή τη ροή που εισέρχεται. (Από [36])

Η Εικόνα 2.12 δείχνει τα τμήματα 9 ανεξάρτητων συνιστωσών (IC's) $s_i(t), i = 1, \dots, 9$ που βρέθηκαν από τα καταγεγραμμένα δεδομένα μαζί με τα αντίστοιχα πρότυπα δυναμικών[36]. Οι δύο πρώτες ανεξάρτητες συνιστώσες οφείλονται σαφώς στη μυϊκή δραστηριότητα που δημιουργείται από το δάγκωμα. Ο διαχωρισμός τους σε δύο συνιστώσες φαίνεται να ανταποκρίνεται, με βάση τα πρότυπα δυναμικών, σε δύο διαφορετικές ομάδες μυών που ενεργοποιήθηκαν κατά τη διάρκεια της διαδικασίας. Οι IC3 και IC5 δείχνουν τις οριζόντιες κινήσεις των ματιών και το ανοιγοκλείσιμο των βλεφάρων, αντίστοιχα. Η IC4 αντιπροσωπεύει το καρδιακό παράσιτο που εξάγεται καθαρά.

Για να βρεθούν τα υπόλοιπα παράσιτα, τα δεδομένα πέρασαν μέσα από υψιπερατό φίλτρο, με συχνότητα αποκοπής 1Hz. Ακολούθως, βρέθηκε η

ανεξάρτητη συνιστώσα IC8, η οποία δείχνει το παράσιτο που προέρχεται σαφώς από το ψηφιακό ρολόι, που βρίσκεται στην δεξιά πλευρά του μαγνητόμετρου. Η τελευταία ανεξάρτητη συνιστώσα IC9 σχετίζεται με έναν αισθητήρα που παρουσιάζει μεγαλύτερη RMS (ενεργό) τιμή θορύβου από τους άλλους.

Τα αποτελέσματα της Εικόνας 2.12 δείχνουν καθαρά ότι χρησιμοποιώντας την τεχνική ICA και τον αλγόριθμο FastICA, είναι δυνατόν να απομονωθούν τα παράσιτα τόσο της κίνησης των ματιών όσο και του ανοιγοκλεισίματος των βλεφάρων, καθώς και τα καρδιακά μυογραφικά και άλλα παράσιτα από τα σήματα της MEG. Ο αλγόριθμος FastICA είναι ένα ιδιαίτερα κατάλληλο εργαλείο, διότι η απομάκρυνση των παρασίτων είναι μία διαδραστική τεχνική και ο ερευνητής μπορεί ελεύθερα να επιλέξει τον αριθμό των ανεξάρτητων συνιστωσών που θέλει.

Πέρα από τον περιορισμό των παρασίτων, η ICA μπορεί να χρησιμοποιηθεί και για να αποσυντεθούν προκλητικά δυναμικά[37], το οποίο επιτρέπει άμεση πρόσβαση στην υποκείμενη εγκεφαλική λειτουργία, η οποία είναι πιθανό να είναι ιδιαίτερα μεγάλης σημασίας για τη νευροεπιστημονική έρευνα.

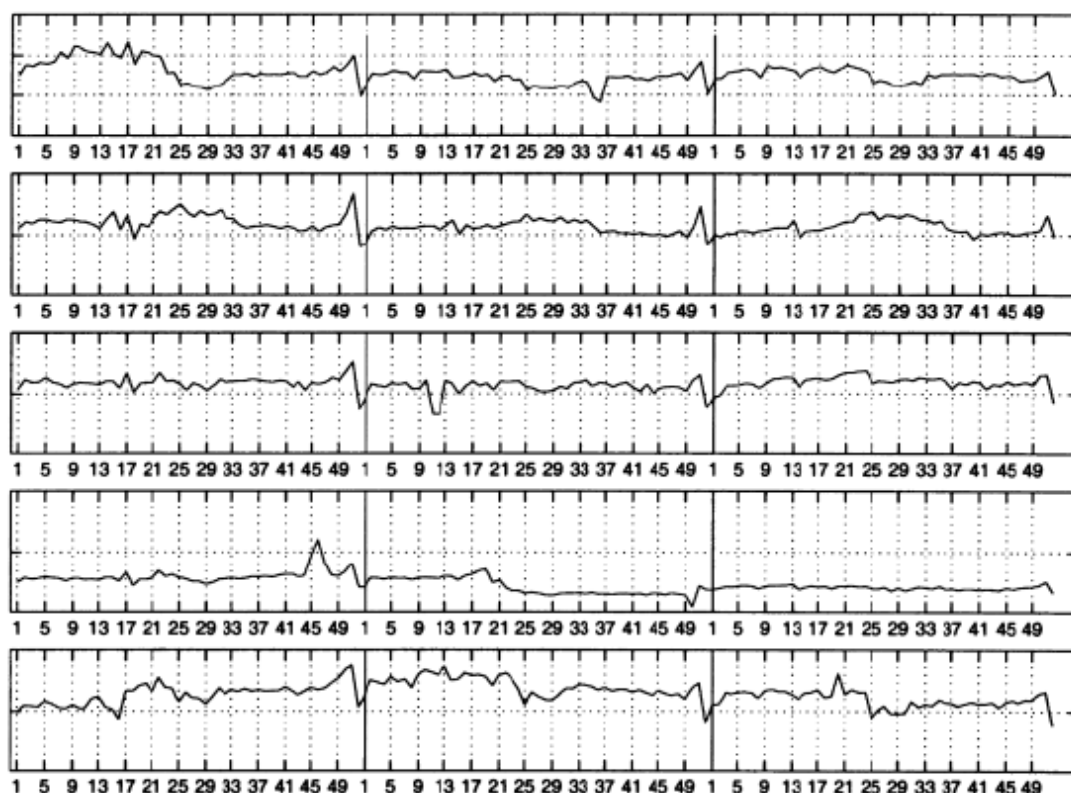
2.7.2 Εύρεση κρυμμένων παραγόντων σε οικονομικά δεδομένα

Μία εναλλακτική λύση είναι η δοκιμή της ICA σε οικονομικά δεδομένα. Υπάρχουν πολλές περιπτώσεις στο πεδίο αυτής της εφαρμογής όπου είναι διαθέσιμες παράλληλες χρονοσειρές, όπως συναλλαγματικές ισοτιμίες ή ημερήσια απόδοση των αποθεμάτων, που μπορεί να έχουν ορισμένους κοινούς προσδιοριστικούς παράγοντες. Η ICA μπορεί να αποκαλύψει κάποιους μηχανισμούς οδήγησης που διαφορετικά παραμένουν κρυμμένοι. Σε μελέτη ενός χαρτοφυλακίου μετοχών [38], διαπιστώθηκε ότι η ICA αποτελεί συμπληρωματικό εργαλείο για την PCA, επιτρέποντας στην υποκείμενη δομή των δεδομένων να είναι πιο εύκολα παρατηρήσιμη.

Η ICA εφαρμόστηκε σε ένα άλλο πρόβλημα [39]: στις ταμειακές ροές από πολλά καταστήματα που ανήκουν στην ίδια αλυσίδα καταστημάτων λιανικής πώλησης. Σκοπός ήταν η εύρεση των κοινών σε όλα τα καταστήματα θεμελιωδών παραγόντων που επηρεάζουν τα δεδομένα των ταμειακών ροών. Έτσι, θα μπορούσε να αναλυθεί η επίδραση των ειδικών παραγόντων για κάθε συγκεκριμένο κατάστημα στις ταμειακές ροές, δηλαδή, η επίδραση των δράσεων που λαμβάνονται στα ξεχωριστά καταστήματα και στο τοπικό τους περιβάλλον.

Η υπόθεση ότι υπάρχουν κάποιες υποκείμενες ανεξάρτητες συνιστώσες σε αυτή τη συγκεκριμένη εφαρμογή μπορεί να είναι ρεαλιστική. Για παράδειγμα, παράγοντες όπως οι εποχιακές διακυμάνσεις λόγω των διακοπών

και οι ετήσιες διακυμάνσεις, και παράγοντες που έχουν ξαφνικές επιπτώσεις στην αγοραστική δύναμη των πελατών όπως αλλαγές στην τιμή διαφόρων εμπορευμάτων, αναμένεται να έχουν επίδραση σε όλα τα καταστήματα λιανικής πώλησης, και τέτοιοι παράγοντες μπορούμε να υποθέσουμε ότι είναι σχεδόν ανεξάρτητοι μεταξύ τους. Ωστόσο, ανάλογα με την πολιτική και τις δεξιότητες του εκάστοτε διευθυντή, π.χ. διαφημιστικές προσπάθειες, η επίδραση των παραγόντων στην ταμειακή ροή συγκεκριμένων καταστημάτων λιανικής πώλησης είναι ελαφρώς διαφορετική. Με την ICA, είναι δυνατόν να απομονωθούν τόσο οι υποκείμενοι παράγοντες όσο και οι στάθμες της επίδρασης, καθιστώντας έτσι δυνατή την ομαδοποίηση των καταστημάτων με βάση τις διαχειριστικές τους πολιτικές χρησιμοποιώντας μόνο τις χρονοσειρές δεδομένων των ταμειακών ροών.

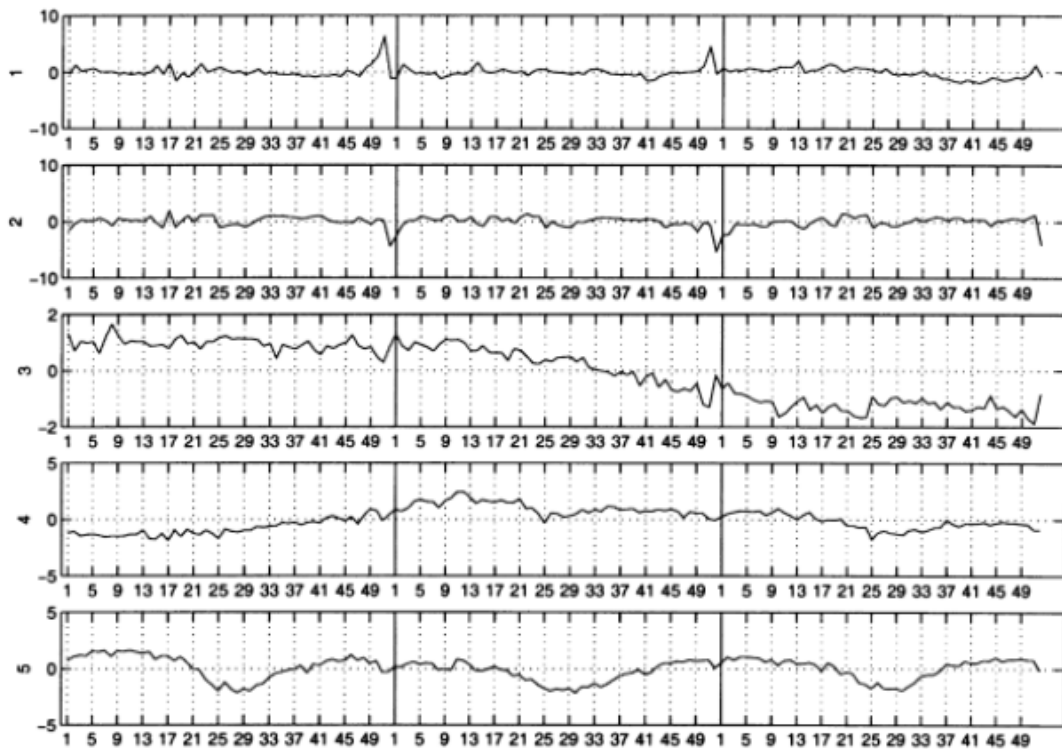


Εικόνα 2.13 Πέντε δείγματα των αρχικών χρονικών σειρών των ταμειακών ροών (η μέση τιμή έχει απομακρυνθεί με κανονικοποίηση στη μοναδιαία τυπική απόκλιση). Οριζόντιος άξονας: χρόνος σε εβδομάδες. (Από [39]).

Τα δεδομένα συνίστατο από την εβδομαδιαία ταμειακή ροή σε 40 καταστήματα που ανήκουν στην ίδια αλυσίδα καταστημάτων λιανικής πώλησης οι μετρήσεις ταμειακών ροών καλύπτουν 140 εβδομάδες. Μερικά παραδείγματα των αρχικών δεδομένων $x_i(t)$ παρουσιάζονται στην Εικόνα 2.13.

Η προ-αποσυσχέτιση (prewhitening) έγινε έτσι ώστε τα αρχικά διανύσματα των σημάτων να έχουν προβολές στον υποχώρο που καλύπτεται από τις πέντε πρώτες κύριες συνιστώσες τους και οι διασπορές να κανονικοποιηθούν στο 1. Έτσι, η διάσταση του χώρου του σήματος μειώθηκε από 40 σε 5. Χρησιμοποιώντας τον αλγόριθμο FastICA, υπολογίζονται πέντε ανεξάρτητες συνιστώσες $s_i(t)$, $i = 1, \dots, 5$. Όπως απεικονίζεται στην Εικόνα 14, ο αλγόριθμος FastICA έχει βρει αρκετούς σαφώς διαφορετικούς θεμελιώδεις παράγοντες κρυμμένους στα αρχικά δεδομένα.

Οι παράγοντες έχουν διαφορετικές ερμηνείες. Οι δύο πρώτοι παράγοντες ακολουθούν τις ξαφνικές αλλαγές που προκαλούνται από διακοπές, κλπ.. Το πιο χαρακτηριστικό παράδειγμα είναι η περίοδος των Χριστουγέννων. Ο παράγοντας στην τελευταία γραμμή, από την άλλη πλευρά, αντανακλά τη βραδύτερη εποχιακή διακύμανση, με την επίδραση των καλοκαιρινών σαφώς ορατή. Ο παράγοντας στην Τρίτη γραμμή θα μπορούσε να αντιπροσωπεύει μία ακόμη πιο αργή μεταβολή, κάτι που μοιάζει με μία τάση. Ο τελευταίος παράγοντας, στην τέταρτη σειρά, είναι διαφορετικός από τους άλλους: θα μπορούσε να είναι επειδή αυτός ο παράγοντας ακολουθεί κυρίως τη σχετική ανταγωνιστική θέση της αλυσίδας των καταστημάτων σε σχέση με τους ανταγωνιστές, αλλά και άλλες ερμηνείες είναι επίσης δυνατές.



Εικόνα 2.14 Πέντε ανεξάρτητες συνιστώσες ή θεμελιώδεις παράγοντες που βρέθηκαν από τα δεδομένα ταμειακών ροών. (Από [39]).

2.7.3 Μείωση θορύβου στις φυσικές εικόνες (natural images)

Το τρίτο παράδειγμα αφορά στην εύρεση φίλτρων ICA για τις φυσικές εικόνες και βασιζόμενο στην αποσύνθεση ICA, στην απομάκρυνση του θορύβου από εικόνες αλλοιωμένες με προσθετικό Γκαουσιανό θόρυβο.

Χρησιμοποιήθηκε μία σειρά από ψηφιακοποιημένες φυσικές εικόνες. Συμβολίζουμε το διάνυσμα διαβαθμίσεων του γκρι των εικονοστοιχείων (pixels) στο παράθυρο εικόνας (image window) με $\mathbf{x}$. Σημειώνουμε ότι, σε αντίθεση με τις δύο άλλες εφαρμογές στις προηγούμενες ενότητες, αυτή τη φορά δεν εξετάζουμε χρονολογικές σειρές με πολλαπλές τιμές ή εικόνες που αλλάζουν με την πάροδο του χρόνου· αντίθετα, τα στοιχεία του $\mathbf{x}$ κατατάσσονται με βάση τη θέση στο παράθυρο εικόνας ή στο επίγραμμα. Τα παράθυρα- δείγματα ελήφθησαν σε τυχαίες θέσεις. Η δισδιάστατη δομή των παραθύρων δεν έχει καμία σημασία: χρησιμοποιήθηκε σάρωση ανά γραμμή για τη μετατροπή ενός τετραγωνικού παραθύρου εικόνας σε διάνυσμα με τιμές εικονοστοιχείων. Οι ανεξάρτητες συνιστώσες τέτοιων παραθύρων εικόνων παρουσιάζονται στην Εικόνα 5. Κάθε παράθυρο σε αυτή την εικόνα αντιστοιχεί σε μία από τις στήλες $\mathbf{a}_i$ του πίνακα μίξης $\mathbf{A}$. Έτσι, ένα παρατηρούμενο παράθυρο εικόνας είναι μία υπέρθεση αυτών των παραθύρων όπως στην εξίσωση 2.5, με ανεξάρτητους συντελεστές [23], [40].

Τώρα, ας υποθέσουμε ένα μοντέλο εικόνας με θόρυβο:

$$\mathbf{z} = \mathbf{x} + \mathbf{n} \quad (2.49)$$

όπου $\mathbf{n}$ είναι ο ασυσχέτιστος θόρυβος με στοιχεία που κατατάσσονται στο παράθυρο εικόνας με τον ίδιο τρόπο όπως στον $\mathbf{x}$, και $\mathbf{z}$ είναι το μετρούμενο παράθυρο εικόνας αλλοιωμένο με τον θόρυβο. Ας υποθέσουμε επιπλέον ότι το $\mathbf{n}$ είναι Γκαουσιανό και το $\mathbf{x}$ είναι μη-Γκαουσιανό. Υπάρχουν πολλοί τρόποι για να καθαρίσουμε τον θόρυβο· ένα παράδειγμα είναι να κάνουμε ένα μετασχηματισμό στη χωρική συχνότητα με DFT (discrete Fourier transform- διακριτός μετασχηματισμός Fourier), βαθυπερατό φιλτράρισμα, και να επιστρέψουμε στο χώρο της εικόνας με IDFT (inverse discrete Fourier transform- αντίστροφος διακριτός μετασχηματισμός Fourier) [12]. Αυτό, ωστόσο, δεν είναι πολύ αποτελεσματικό. Μία καλύτερη μέθοδος Συρρίκνωσης Κυματίων (Wavelet Shrinkage method) [41] στην οποία χρησιμοποιείται ένας μετασχηματισμός που βασίζεται σε κυμάτια, ή μέθοδοι που βασίζονται σε μέσο φιλτράρισμα (median filtering)[12]. Ωστόσο, καμία από αυτές τις μεθόδους δεν υπερτερεί.

Μία άλλη μέθοδος που διέπεται από τις αρχές της στατιστικής και ονομάζεται Sparse Code Shrinkage [42], είναι πολύ στενά συνδεδεμένη με την

ανάλυση ανεξάρτητων συνιστωσών. Εν συντομία, αν μοντελοποιήσουμε την πυκνότητα πιθανότητας του $\mathbf{x}$ με ICA, και υποθέσουμε ότι το $\mathbf{n}$ είναι Γκαουσιανό, τότε η λύση Μέγιστης Πιθανοφάνειας (ML) για το $\mathbf{x}$ δοθέντων των μετρήσεων $\mathbf{z}$ μπορεί να αναπτυχθεί στο μοντέλο σήματος (εξίσωση 2.49).

Η λύση ML μπορεί εύκολα να υπολογιστεί, έστω και κατά προσέγγιση, χρησιμοποιώντας μία αποσύνθεση που αποτελεί ορθογωνιοποιημένη έκδοση της ICA. Ο μετασχηματισμός δίνεται από τη σχέση:

$$\mathbf{Wz} = \mathbf{Wx} + \mathbf{Wn} = \mathbf{s} + \mathbf{Wn} \quad (2.50)$$

όπου $\mathbf{W}$ είναι εδώ ένας ορθογώνιος πίνακας που είναι η καλύτερη ορθογώνια προσέγγιση του αντίστροφου πίνακα μίξης της ICA. Ο όρος για τον θόρυβο $\mathbf{Wn}$ είναι Γκαουσιανός και λευκός. Με έναν κατάλληλα επιλεγμένο ορθογώνιο μετασχηματισμό $\mathbf{W}$, ωστόσο, η πυκνότητα πιθανότητας του $\mathbf{Wx}=\mathbf{s}$ γίνεται έντονα μη-Γκαουσιανή με υψηλή θετική κύρτωση. Αυτό εξαρτάται, φυσικά, από τα αρχικά σήματα $\mathbf{x}$, καθώς υποθέτουμε στην πραγματικότητα ότι υπάρχει ένα μοντέλο $\mathbf{x}=\mathbf{W}^T\mathbf{s}$ για το σήμα, τέτοιο τα «σήματα πηγής» (ή στοιχεία του $\mathbf{s}$) να έχουν θετική κυρτωτική πυκνότητα πιθανότητας, όπου σε αυτή την περίπτωση ο μετασχηματισμός της ICA δίνει έντονα υπερ-Γκαουσιανές συνιστώσες. Αυτό φαίνεται να ισχύει τουλάχιστον για πράθυρα εικόνων φυσικών σκηνών[43].

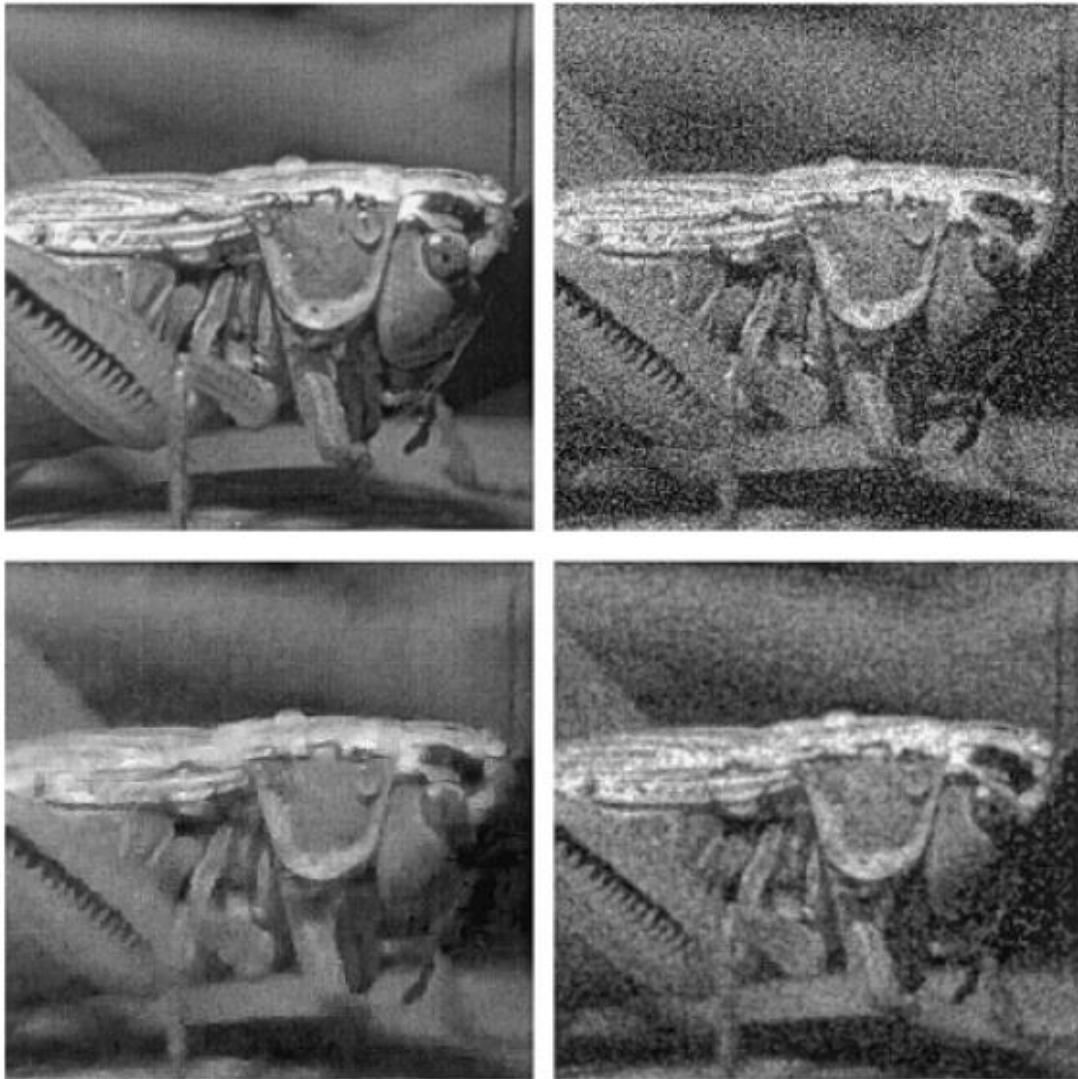
Υποθέτοντας [42] μία πυκνότητα πιθανότητας που ακολουθεί την κατανομή Laplace για τα s_i , η λύση ML για τα s_i δίνεται από μία «συνάρτηση συρρίκνωσης» (“shrinkage function”) $\hat{s}_i=g([\mathbf{Wz}]_i)$, ή σε μορφή διανύσματος, $\hat{\mathbf{s}}=g([\mathbf{Wz}]_i)$ [42]. Η συνάρτηση $g(\cdot)$ έχει χαρακτηριστικό σχήμα: είναι μηδέν κοντά στην αρχή των αξόνων και στη συνέχεια γραμμική μετά από μία συγκεκριμένη τιμή που εξαρτάται από τις παραμέτρους της πυκνότητας πιθανότητας που ακολουθεί την κατανομή Laplace και της πυκνότητας πιθανότητας του Γκαουσιανού θορύβου. Υποθέτοντας άλλες μορφές για τις πυκνότητες πιθανότητας, άλλες βέλτιστες συναρτήσεις συρρίκνωσης μπορούν να προκύψουν [42].

Στη μέθοδο Sparse Code Shrinkage, η πράξη της συρρίκνωσης εφαρμόζεται στον περιστρεμμένο χώρο, μετά από την οποία η εκτίμηση για το σήμα στον αρχικό χώρο δίνεται από αντίστροφη περιστροφή:

$$\hat{\mathbf{x}} = \mathbf{W}^T\hat{\mathbf{s}} = \mathbf{W}^Tg(\mathbf{Wz}) \quad (2.51)$$

Έτσι, έχουμε τον υπολογισμό της Μέγιστης Πιθανοφάνειας για το παράθυρο εικόνας στο οποίο μεγάλο μέρος του θορύβου έχει αφαιρεθεί.

Ο τελεστής περιστροφής $\mathbf{W}$ είναι τέτοιος που η αραιότητα των συνιστωσών $\mathbf{s}=\mathbf{Wx}$ αυξάνεται. Ο εν λόγω τελεστής μπορεί να εκπαιδευτεί με μία τροποποίηση του αλγορίθμου FastICA [42].



Εικόνα 15 Ένα πείραμα στην απομάκρυνση θορύβου. Επάνω αριστερά: η αρχική εικόνα. Επάνω δεξιά: η αρχική εικόνα αλλοιωμένη με θόρυβο (το επίπεδο θορύβου είναι 50%). Κάτω αριστερά: η ανακτημένη εικόνα μετά την εφαρμογή της μεθόδου *Sparse Code Shrinkage*. Κάτω δεξιά: για σύγκριση, μία εικόνα φιλτραρισμένη με φίλτρο *wiener*.

Ένα αποτέλεσμα καθαρισμού θορύβου φαίνεται στην Εικόνα 15. Εμφανίζονται μία χωρίς θόρυβο εικόνα και μία έκδοση με θόρυβο, στην οποία το επίπεδο θορύβου είναι το 50% του επιπέδου του σήματος. Τα αποτελέσματα της μεθόδου *Sparse Code Shrinkage* και της κλασικής μεθόδου φιλτραρίσματος Wiener είναι δεδομένα υποδεικνύοντας ότι η μέθοδος *Sparse Code Shrinkage* μπορεί να είναι μία υποσχόμενη προσέγγιση. Ο θόρυβος έχει μειωθεί χωρίς να αλλοιώνονται οι ακμές ή άλλα αιχμηρά χαρακτηριστικά, τόσο πολύ όσο στο φιλτράρισμα Wiener. Αυτό οφείλεται σε μεγάλο βαθμό στην έντονα μη γραμμική φύση του τελεστή συρρίκνωσης που είναι βέλτιστα προσαρμοσμένος στην εγγενή στατιστική των φυσικών εικόνων.

2.7.4 Τηλεπικοινωνίες

Άλλη μία ανερχόμενη εφαρμογή της μεθόδου ICA με μεγάλες δυνατότητες είναι οι τηλεπικοινωνίες. Ένα παράδειγμα μιας πραγματικής εφαρμογής τηλεπικοινωνιών όπου οι τεχνικές τυφλού διαχωρισμού είναι χρήσιμες είναι ο διαχωρισμός του σήματος του ίδιου του χρήστη από τα σήματα άλλων χρηστών που παρεμβάλλονται στις CMDA (Code-Division Multiple Access-πολλαπλή πρόσβαση διαίρεσης κώδικα) κινητές επικοινωνίες [44]. Αυτό το πρόβλημα είναι κατά το ήμισυ «τυφλό», υπό την έννοια ότι είναι διαθέσιμη συγκεκριμένη προσθετική προγενέστερη πληροφορία στο μοντέλο δεδομένων CDMA. Αλλά ο αριθμός των παραμέτρων που πρέπει να υπολογιστούν είναι συχνά τόσο μεγάλος που οι κατάλληλες τεχνικές τυφλού διαχωρισμού πηγαιών σημάτων που λαμβάνουν υπόψη τη διαθέσιμη προγενέστερη γνώση παρέχουν μία σαφή βελτίωση της απόδοσης από τις παραδοσιακότερες τεχνικές εκτίμησης [44].

Κεφάλαιο 3

Εφαρμογές της μεθόδου ICA σε πρότυπα σήματα και σήματα φωνής

3.1 Εισαγωγή

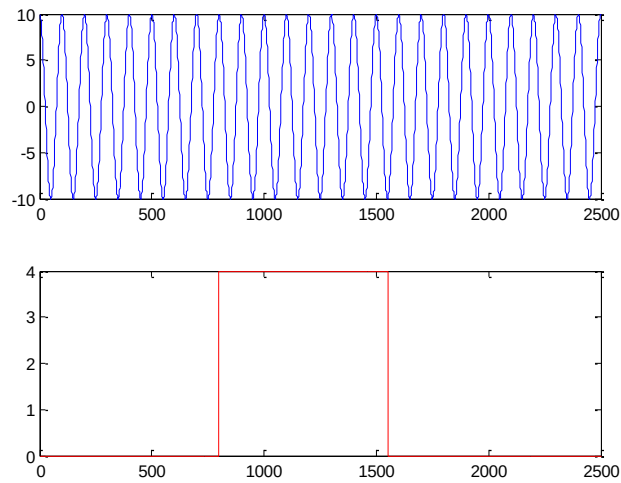
Σε αυτό το κεφάλαιο θα πραγματοποιηθούν εφαρμογές με τη μέθοδο ICA σε πρότυπα σήματα και σήματα φωνητικών σημάτων (wave). Η μέθοδος αυτή εκτελείται με τη χρήση του αλγορίθμου `fastICA`, ο οποίος δίνει τη δυνατότητα διαχωρισμού των πηγών από τα σήματα μίξης. Σε κάθε εφαρμογή υπάρχει ένας αριθμός ομιλητών οι οποίοι μιλούν ταυτόχρονα, έτσι προκύπτουν τα σήματα μίξης και ο σκοπός του προβλήματος είναι ο διαχωρισμός της φωνής καθενός. Η μέθοδος ICA μας δίνει τη δυνατότητα επίλυσης αυτού του προβλήματος και όπως φαίνεται παρακάτω, το αποδεικνύουμε ηχητικά καθώς και γραφικά.

Για να λειτουργήσει ο αλγόριθμος `fastICA` απαιτείται η συνάρτηση ICA, δηλαδή ένα m-file που περιλαμβάνει τον κώδικα `fastICA` και μέσα στο ίδιο m-file υπολογίζεται και ο κώδικας ICA. Αυτό το m-file βρίσκεται στο παράρτημα της εργασίας, όπου εξηγείται ο τρόπος λειτουργίας του.

3.2 Εφαρμογές σε πρότυπα σήματα

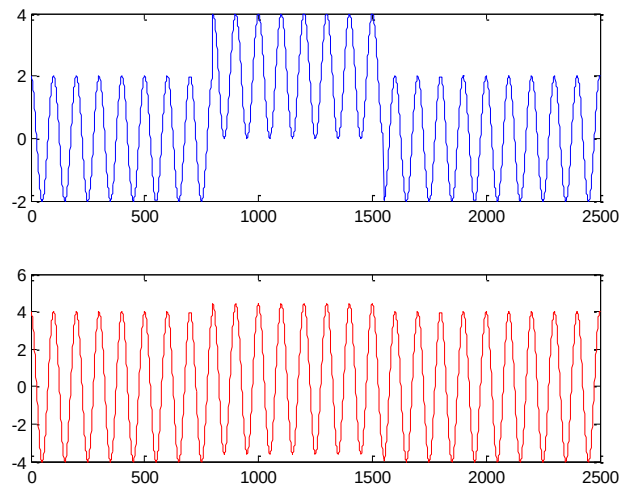
Σε αυτή την ενότητα θα εφαρμόσουμε τον αλγόριθμο `fastICA` σε πρότυπα σήματα (cos, παλμό) , με σκοπό το διαχωρισμό των πηγών από τα σήματα μίξης.

Κατασκευάζουμε δύο σήματα, ένα ημιτονοειδές x_1 και έναν παλμό x_2 :



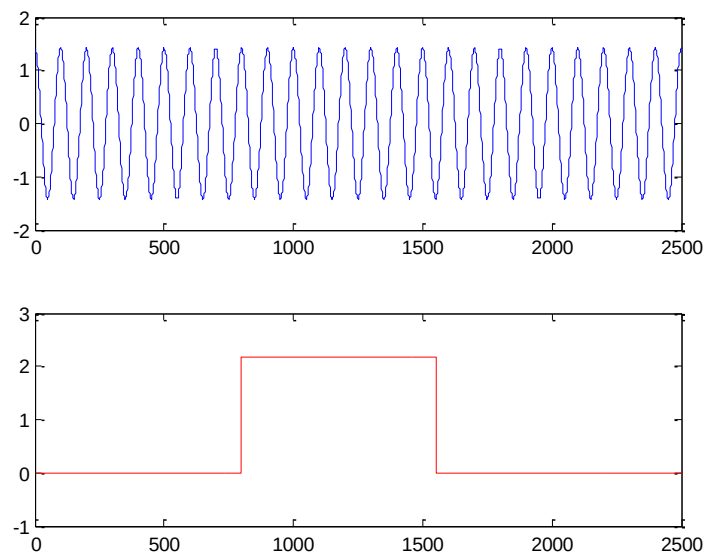
Σχήμα 3.16 Τα αρχικά σήματα x_1, x_2 (red) που κατασκευάστηκαν.

και έπειτα δημιουργούμε δύο νέα σήματα y_1, y_2 , τα οποία είναι συνδυασμός των αρχικών:



Σχήμα 3.2 Η μίξη των σημάτων $y_1 = \alpha_{11}x_1 + \alpha_{12}x_2$ και $y_2 = \alpha_{21}x_1 + \alpha_{22}x_2$

Μετά την εφαρμογή του αλγορίθμου fastICA στα y_1, y_2 προέκυψαν οι εκτιμήσεις c_1, c_2 των αρχικών σημάτων x_1, x_2 :



Σχήμα 3.3 Οι εκτιμήσεις c_1, c_2 (red) των αρχικών σημάτων x_1, x_2 μετά την εφαρμογή της ICA.

Όπως φαίνεται από τα παραπάνω αποτελέσματα, τα σήματα που προέκυψαν μετά την εφαρμογή της ICA είναι πολύ κοντά στα αρχικά σήματα, έχουν την ίδια μορφή αλλά μικρότερο πλάτος.

Ο αλγόριθμος που χρησιμοποιήθηκε είναι ο ακόλουθος:

```
%δημιουργία σήματος x1 (ημίτονο με πλάτος=10 και συχνότητα
%δειγματοληψίας=1KHz)
for(i=1:2500)
x1(i)=10*cos(2*pi*10*i/1000);
end

%δημιουργία σήματος x2 (παλμός με πλάτος=4)
x2=zeros(1,2500);
for(j=800:1550)
x2(1,j)=4;
end

figure;
subplot(2,1,1); plot(x1); % plot x1
subplot(2,1,2); plot(x2, 'r'); % plot x2

%αρχικές τιμές βαρών
a1=0.2; a2=0.5; a3=0.4; a4=0.1;

%μίξη αρχικών σημάτων x1 και x2
for(n=1:2500)
```

```

y1(n)=a1*x1(n)+a2*x2(n);
y2(n)=a3*x1(n)+a4*x2(n);
end

%σχεδιάζουμε τη μίξη των σημάτων
figure;
subplot(2,1,1); plot(y1);
subplot(2,1,2); plot(y2, 'r');

figure;
c = fastica([y1;y2]); % εφαρμογή του αλγορίθμου fastICA
subplot(2,1,1); plot(c(1,:));
subplot(2,1,2); plot(c(2,:), 'r');

```

3.3 Εφαρμογή σε σήματα φωνής εξαρτημένα από τις αποστάσεις των ομιλητών χρησιμοποιώντας τον αλγόριθμο fastICA.

Σε αυτή την εφαρμογή χρησιμοποιούμε τις φωνές δύο ομιλητών, με τα ανάλογα βάρη a_{ij} (αποστάσεις των ομιλητών από τα μικρόφωνα) και αφού δημιουργήσουμε τα σήματα μίξης με τη βοήθεια του αλγορίθμου fastICA κάνουμε διαχωρισμό των σημάτων των πηγών (ανεξάρτητες συνιστώσες). Το m-file που απαιτείται για αυτή την εφαρμογή, παρουσιάζεται αναλυτικά σύμφωνα με τον παρακάτω κώδικα.

```

% τα σήματα των πηγών  $s_i$ 
[temp,f]=wavread('voice1.wav'); % φορτώνουμε τη φωνή1
temp=temp';
s=temp(1,:); % το σήμα της πηγής  $s_1$ 
N=length(s);
sound(s(1,:),f); %ακούμε τη φωνή1

[temp,f]=wavread('voice2.wav'); % φορτώνουμε τη φωνή2
temp=temp';
s(2,:)=temp(1,1:N); % το σήμα της πηγής  $s_2$ 
sound(s(2,:),f); %ακούμε τη φωνή2

A=[0.5 0.2;0.3 0.6]; %δημιουργούμε τον πίνακα μίξης
x=A*s; %διαδικασία μίξης

c=fastica([x(1,:);x(2,:)]); % εφαρμογή αλγορίθμου fastICA

%ακούμε τη μίξη των δύο σημάτων
sound(x(1,:),f);
sound(x(2,:),f);

```

%ακούμε τις ανεξάρτητες συνιστώσες που επιστρέφει ο αλγόριθμος fastICA

```
sound(c(1,:),f);
sound(c(2,:),f);
```

%παρουσιάζουμε τα σήματα των πηγών

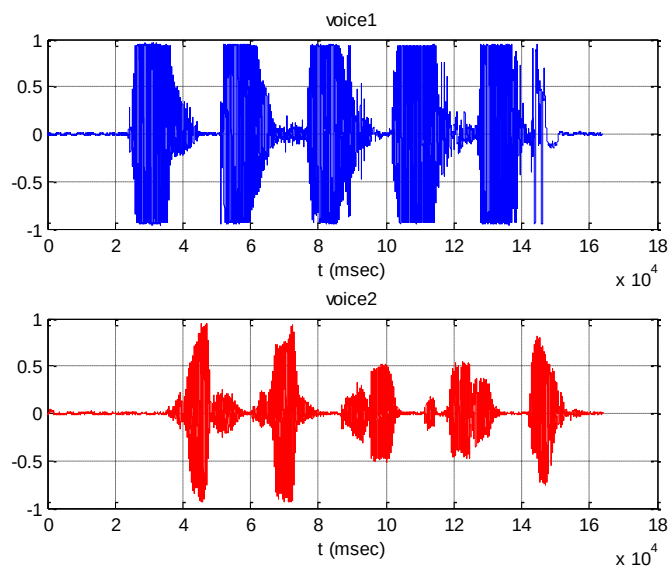
```
figure;
subplot(2,1,1); plot(s(1,:)),grid on, title('voice1'), xlabel('t (msec)'); % plot
s1
subplot(2,1,2); plot(s(2,:), 'r'),grid on, title('voice2'), xlabel('t (msec)'); % plot
s2
```

%παρουσιάζουμε τις ανεξάρτητες συνιστώσες που επιστρέφει ο αλγόριθμος fastICA

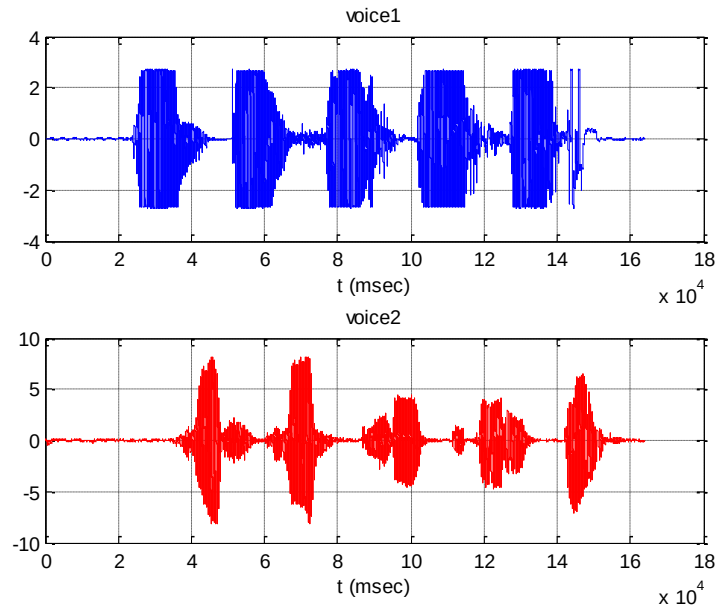
```
subplot(2,1,1); plot(c(1,:)), grid on,title('voice1'), xlabel('t (msec)');
subplot(2,1,2); plot(c(2,:), 'r'), grid on, title('voice2'), xlabel('t (msec)');
```

Στη συνέχεια θα περιγράψουμε αναλυτικά τα αποτελέσματα που προκύπτουν μετά την εφαρμογή του αλγορίθμου fastICA.

Στο παρακάτω σχήμα φαίνονται τα αρχικά σήματα, καθώς και τα αποτελέσματα που επιστρέφει ο αλγόριθμος fastICA.



(α)



(β)

Σχήμα 3.4 (α) Τα αρχικά σήματα. (β) Τα αποτελέσματα που επιστρέφει ο αλγόριθμος *fastICA*.

Παρατηρούμε ότι για κάθε ομιλητή, μετά την εφαρμογή του αλγορίθμου στα σήματα μίξης τα αποτελέσματα είναι ίδια με τα αρχικά αλλά με μεγαλύτερο πλάτος (αλλάζει η σειρά των πηγών αλλά αυτό δεν επηρεάζει τα αποτελέσματα). Άρα η μέθοδος ICA σε αυτή την εφαρμογή των ηχητικών σημάτων δίνει σωστά αποτελέσματα, με τη βοήθεια του αλγορίθμου *fastICA*.

3.4 Εφαρμογή του *fastICA* σε ένα σήμα φωνής και μια μελωδία.

Σε αυτή την εφαρμογή χρησιμοποιούμε τη φωνή ενός ομιλητή και μια μελωδία από κρουστά, με τα ανάλογα βάρη a_{ij} (αποστάσεις από τα μικρόφωνα) και αφού δημιουργήσουμε τα σήματα μίξης με τη βοήθεια του αλγορίθμου *fastICA* κάνουμε διαχωρισμό των σημάτων των πηγών (ανεξάρτητες συνιστώσες). Το m-file που απαιτείται για αυτή την εφαρμογή, παρουσιάζεται αναλυτικά σύμφωνα με τον παρακάτω κώδικα.

```
[temp,f]=wavread('voice.wav'); %φορτώνουμε τη φωνή
temp=temp';
s=temp(1,:); %το σήμα της πηγής  $s_1$ 
N=length(s);
sound(s(1,:),f); %ακούμε τη φωνή1
```

```
[temp,f]=wavread('sound.wav'); %φορτώνουμε τη μελωδία
temp=temp';
```

```
s(2,:)=temp(1,1:N); %σήμα πηγής2
sound(s(2,:),f); %ακούμε τη φωνή2
```

```
A=[0.4 0.2;0.3 0.7]; %δημιουργούμε τον πίνακα μίξης
x=A*s; %διαδικασία μίξης
```

```
c=fastica([x(1,:);x(2,:)]); % εφαρμογή αλγορίθμου fastICA
```

```
%ακούμε τη μίξη των δύο σημάτων
sound(x(1,:),f);
sound(x(2,:),f);
```

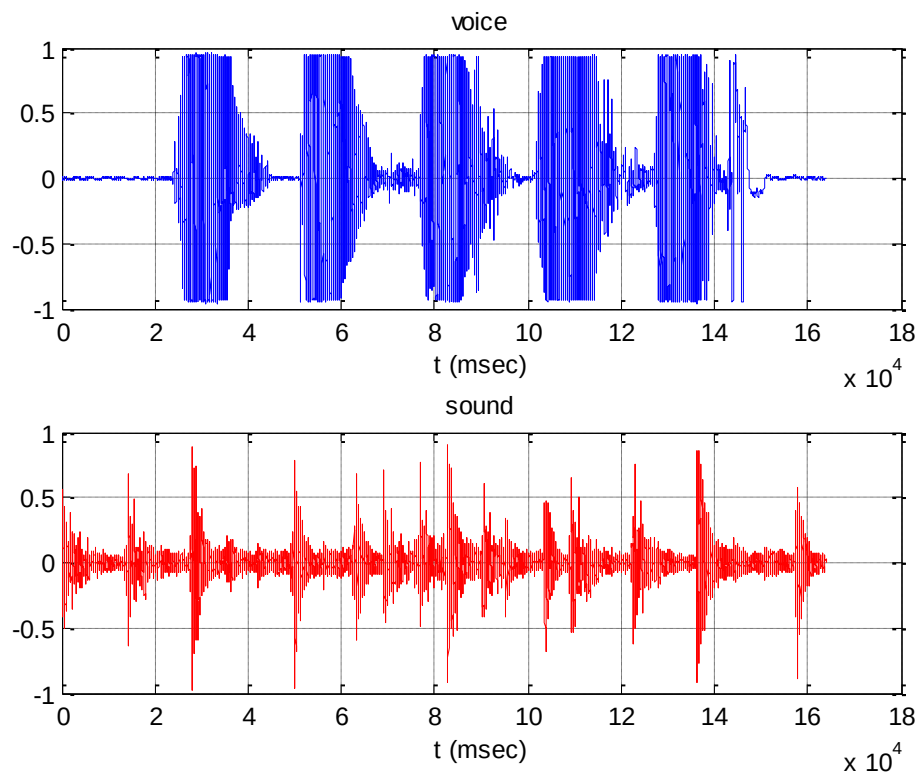
```
%ακούμε τις ανεξάρτητες συνιστώσες που επιστρέφει ο αλγόριθμος fastICA
sound(c(1,:),f);
sound(c(2,:),f);
```

```
%παρουσιάζουμε τα σήματα των πηγών
figure;
subplot(2,1,1); plot(s(1,:)),grid on, title('voice'), xlabel('t (msec)'); % plot s1
subplot(2,1,2); plot(s(2,:), 'r'),grid on, title('sound'), xlabel('t (msec)'); % plot
s2
```

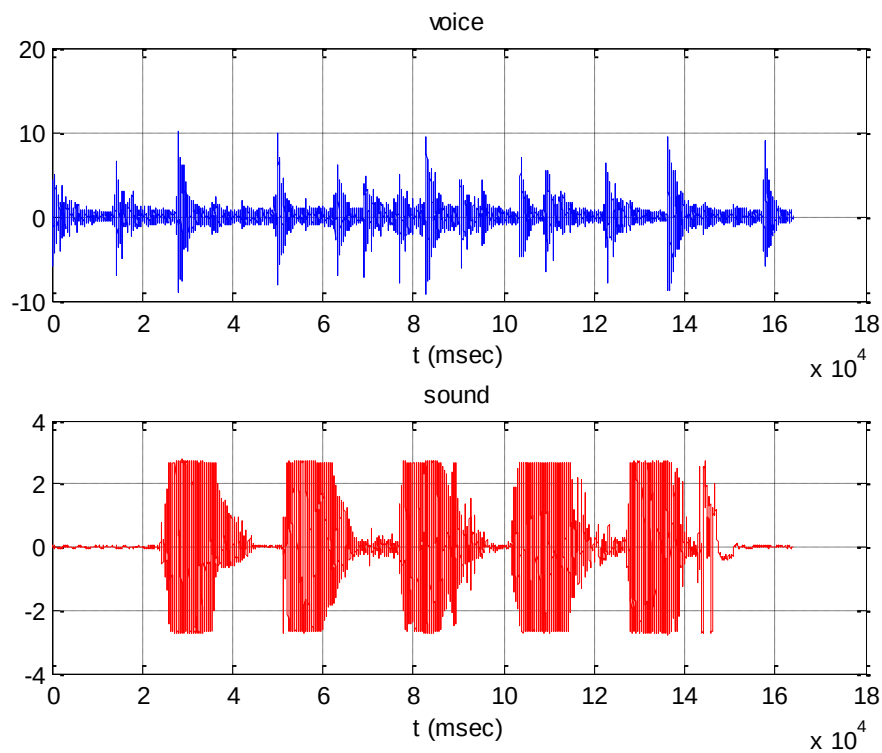
```
%παρουσιάζουμε τις ανεξάρτητες συνιστώσες που επιστρέφει ο αλγόριθμος
fastICA
subplot(2,1,1); plot(c(1,:)), grid on,title('voice'), xlabel('t (msec)');
subplot(2,1,2); plot(c(2,:), 'r'), grid on, title('sound'), xlabel('t (msec)');
```

Παρακάτω περιγράφουμε αναλυτικά τα αποτελέσματα που προκύπτουν μετά την εφαρμογή του αλγορίθμου fastICA.

Στο παρακάτω σχήμα φαίνονται τα αρχικά σήματα, καθώς και τα αποτελέσματα που επιστρέφει ο αλγόριθμος fastICA.



(α)



(β)

Σχήμα 3.5 (α) Τα αρχικά σήματα. (β) Τα αποτελέσματα που επιστρέφει ο αλγόριθμος fastICA.

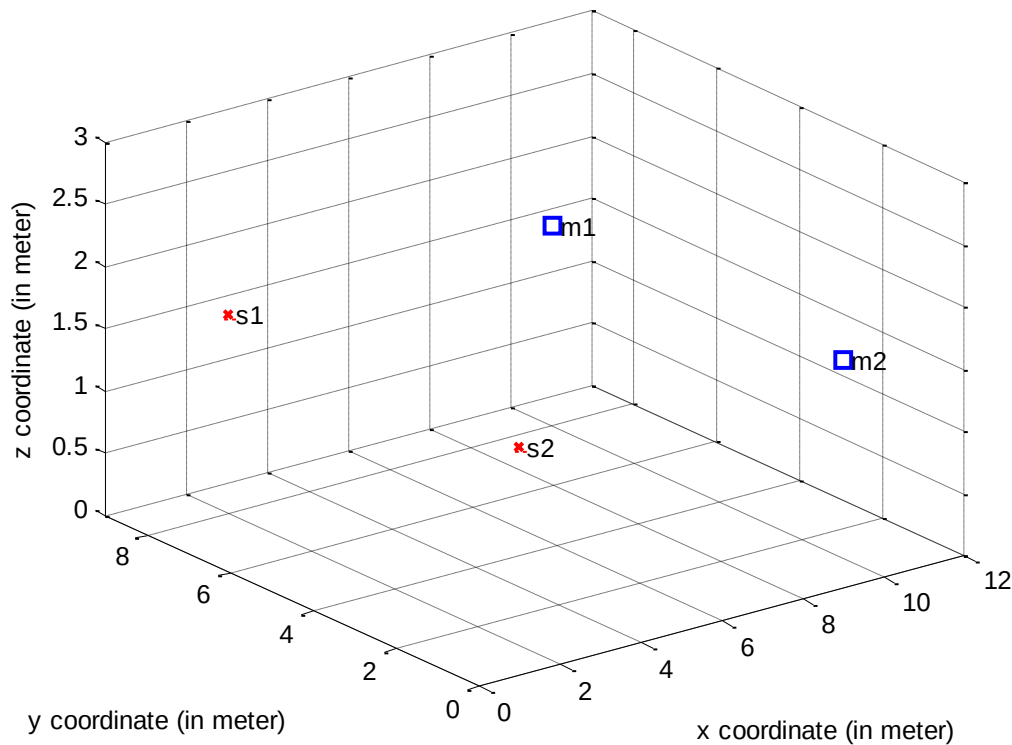
Παρατηρούμε ότι μετά την εφαρμογή του αλγορίθμου στα σήματα μίξης στο σήμα του ομιλητή ακούγεται σε πολύ μικρή ένταση το σήμα της μελωδίας. Αντίθετα το σήμα της μελωδίας που παίρνουμε μετά τον διαχωρισμό είναι ακριβώς ίδιο με το αρχικό.

3.5 Εφαρμογή του fastICA για το διαχωρισμό σημάτων δύο ομιλητών σε ένα δωμάτιο

Ας υποθέσουμε ότι σε ένα ανηχοϊκό δωμάτιο βρίσκονται δύο ομιλητές (πηγές) και δύο μικρόφωνα. Δίνουμε στο Matlab τις συντεταγμένες του δωματίου καθώς και τις συντεταγμένες των ομιλητών και των μικροφώνων. Στη συνέχεια αφού δημιουργήσουμε τα σήματα μίξης χρησιμοποιώντας τα ανάλογα βάρη a_{ij} κάνουμε διαχωρισμό των σημάτων των πηγών με τη βοήθεια του αλγορίθμου fastICA.

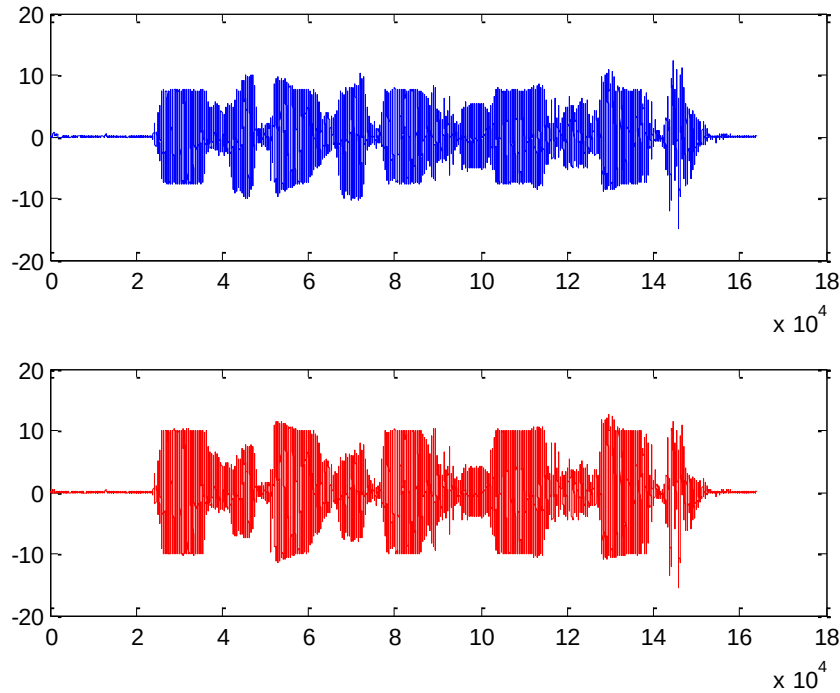
Η απεικόνιση του δωματίου φαίνεται στο παρακάτω σχήμα:

Spatial configuration : localization of the sources and sensors in the room



Σχήμα 3.6 Απεικόνιση των δύο ομιλητών και των δύο μικροφώνων σε ένα δωμάτιο.

Κατά τη μίξη των σημάτων παρατηρούμε την ύπαρξη θορύβου. Η μίξη των σημάτων φαίνεται στο παρακάτω σχήμα:



Σχήμα 3.7 Η μίξη των σημάτων (x_1 και x_2 αντίστοιχα).

Όπως και στις ενότητες που προηγήθηκαν έτσι κι εδώ παρατηρούμε ότι για κάθε σήμα, μετά την εφαρμογή του αλγορίθμου fastICA στα σήματα μίξης τα αποτελέσματα είναι ίδια με τα αρχικά (αλλάζει η σειρά των πηγών αλλά αυτό δεν επηρεάζει τα αποτελέσματα). Οπότε η μέθοδος ICA και σε αυτή την εφαρμογή των ηχητικών σημάτων δίνει σωστά αποτελέσματα.

Κεφάλαιο 4

Το πρόβλημα του Υπερ-ορισμένου διαχωρισμού σημάτων

4.1 Εισαγωγή

Ο τυφλός διαχωρισμός σημάτων είναι μία μέθοδος που μπορεί να χρησιμοποιηθεί αποδοτικά προς την κατεύθυνση της βελτίωσης της ακρίβειας των συστημάτων αναγνώρισης ομιλίας, τόσο στην περίπτωση της στιγμιαίας μίξης σημάτων ομιλίας, όσο και στην γενικότερη και με μεγαλύτερες δυνατότητες πρακτικών εφαρμογών περίπτωση της συνελικτικής μίξης (Κεφάλαιο 5) πολλαπλών ομιλητών σε περιβάλλον πραγματικού δωματίου. Συνήθως ο τυφλός διαχωρισμός σημάτων χρησιμοποιείται στην περίπτωση που έχουμε τον ίδιο αριθμό αισθητήρων και πηγών σημάτων και υποθέτουμε ότι έχουμε ιδανικούς αισθητήρες οι οποίοι δεν προσθέτουν θόρυβο στο σύστημά μας. Η περίπτωση αυτή είναι γνωστή στη βιβλιογραφία ως *πλήρως-ορισμένος διαχωρισμός σημάτων*. Πολύ λίγη έρευνα έχει γίνει για την περίπτωση αισθητήρων που παράγουν θόρυβο. Επιθυμούμε ο θόρυβος να είναι αρκετά μικρός έτσι ώστε να μην επηρεάζει σημαντικά τον αλγόριθμο του τυφλού διαχωρισμού σημάτων. Υποθέτουμε λοιπόν ότι έχουμε χαμηλό SNR στους αισθητήρες και θα δείξουμε ότι ένας πιθανός τρόπος για να βελτιώσουμε την απόδοση του διαχωρισμού είναι να χρησιμοποιήσουμε μεγαλύτερο αριθμό αισθητήρων από τον αριθμό των υπαρχόντων πηγών.

Αυτή η περίπτωση είναι γνωστή ως *υπέρ-ορισμένος τυφλός διαχωρισμός σημάτων*. Με άλλα λόγια ο αριθμός των δεκτών είναι μεγαλύτερος από τον αριθμό των ταυτόχρονων πηγών ομιλίας και εφαρμόζεται τόσο στην περίπτωση της στιγμιαίας όσο και της συνελικτικής μίξης των σημάτων. Η χρήση περισσότερων μικροφώνων αυξάνει την πληροφορία που λαμβάνεται από τα μικρόφωνα. Το γεγονός αυτό θα μπορούσε να βελτιώσει την απόδοση των αλγορίθμων διαχωρισμού και κατά συνέπεια να συμβάλει και στην αύξηση της αξιοπιστίας των χρησιμοποιούμενων συστημάτων αναγνώρισης ομιλίας.

Στο κεφάλαιο αυτό θα μελετήσουμε την επίδραση του πλήθους των χρησιμοποιούμενων δεκτών στη βελτίωση της απόδοσης των συστημάτων αναγνώρισης ομιλίας, παρουσιάζοντας το πρόβλημα του υπερ-ορισμένου τυφλού διαχωρισμού σημάτων (Overdetermined Blind source Separation). Οι μετρήσεις που πραγματοποιήθηκαν αφορούν στην περίπτωση δύο ταυτόχρονων ομιλητών σε ανηχοϊκό θάλαμο. Συγκεκριμένα χρησιμοποιούμε τις φωνές δύο ομιλητών, με τα ανάλογα βάρη a_{ij} (αποστάσεις των ομιλητών από τα μικρόφωνα) και αφού δημιουργήσουμε τα σήματα μίξης, με τη βοήθεια του αλγόριθμου fastICA κάνουμε διαχωρισμό των σημάτων των πηγών (ανεξάρτητες συνιστώσες). Σκοπός των μετρήσεων είναι να μελετήσουμε την επίδραση που έχει η αύξηση του αριθμού των χρησιμοποιούμενων μικροφώνων στην απόδοση ενός συστήματος αναγνώρισης ομιλίας. Στο συμπέρασμα αυτό οδηγούμαστε υπολογίζοντας για όλες τις περιπτώσεις (2 έως 5 μικρόφωνα) το **peak signal-to-noise ratio** ή **PSNR**.

4.2 Υπέρ-ορισμένος διαχωρισμός σε στιγμιαία μίξη

4.2.1 Το μοντέλο της μίξης

Η διαδικασία μίξης περιγράφεται από την παρακάτω εξίσωση:

$$\mathbf{x}(t) = \mathbf{A} \cdot \mathbf{s}(t) \quad (4.1)$$

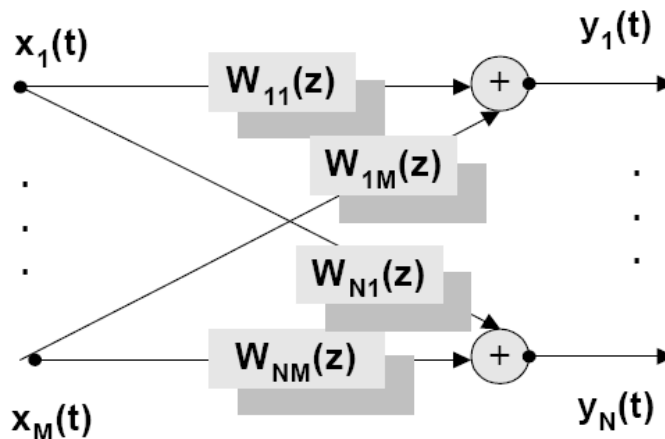
όπου $\mathbf{s}(t) = [s_1(t), \dots, s_N(t)]$, τα καθαρά και στατιστικώς ανεξάρτητα σήματα ομιλίας N ταυτόχρονων ομιλητών και $\mathbf{x}(t) = [x_1(t), \dots, x_M(t)]$ οι παρατηρήσεις που λαμβάνονται από τους M δέκτες. Ο πίνακας $\mathbf{A} \in \mathbb{R}^{M \times N}$ είναι πίνακας μίξης ο οποίος σε αντίθεση με την περίπτωση του πλήρως ορισμένου διαχωρισμού δεν είναι πλέον τετραγωνικός. Η λύση του προβλήματος του υπέρ-ορισμένου διαχωρισμού σημάτων συνίσταται στην εκτίμηση των αρχικών ανεξάρτητων σημάτων $\mathbf{s}(t)$, δεδομένων μόνο των παρατηρήσεων $\mathbf{x}(t)$ και της βασικής υπόθεσης της στατιστικής ανεξαρτησίας των αρχικών σημάτων.

4.2.2 Υπάρχουσες μέθοδοι διαχωρισμού

Το να λύσουμε το πρόβλημα του τυφλού διαχωρισμού σημαίνει ότι πρέπει να βρούμε έναν πίνακα διαχωρισμού $\mathbf{W} \in \mathbb{R}^{M \times N}$ έτσι ώστε το αποτέλεσμα της διαδικασίας διαχωρισμού να είναι:

$$\mathbf{y}(t) = \mathbf{W} \cdot \mathbf{x}(t) \quad (4.2)$$

όπου $\mathbf{y}(t) = [y_1(t), \dots, y_N(t)]$ είναι οι εκτιμήσεις των αρχικών σημάτων $\mathbf{s}(t)$. Το δίκτυο διαχωρισμού που αντιστοιχεί στο γραμμικό μοντέλο φαίνεται στο σχήμα 1.



Σχήμα 4.1 Το δίκτυο διαχωρισμού σημάτων για το υπέρ-ορισμένο πρόβλημα του διαχωρισμού σημάτων.

Εφόσον ο πίνακας διαχωρισμού που χρησιμοποιείται στην περίπτωση του υπέρ-ορισμένου διαχωρισμού δεν είναι τετραγωνικός ο βασικός αλγόριθμος που προκύπτει από το κριτήριο της Μεγιστοποίησης της Πληροφορίας

$$\Delta \mathbf{W} \propto (\mathbf{W}^T)^{-1} - f(\mathbf{y}) \cdot \mathbf{x}^T \quad (4.3)$$

καθώς και η επέκτασή του με χρήση της φυσικής βάθμωσης

$$\Delta \mathbf{W} \propto (\mathbf{I} - f(\mathbf{y}) \cdot \mathbf{y}^T) \cdot \mathbf{W} \quad (4.4)$$

που εφαρμόζονται στην περίπτωση του πλήρως ορισμένου διαχωρισμού δεν μπορούν να χρησιμοποιηθούν αφού δεν μπορούμε να υπολογίσουμε τον αντίστροφο του πίνακα.

Παρόλα αυτά, οι λύσεις που δίνονται μπορούν να επεκταθούν και για την περίπτωση του υπέρ-ορισμένου διαχωρισμού σημάτων [45], αντικαθιστώντας τον αντίστροφο πίνακα που εμφανίζεται στην σχέση με τον ψευδοαντίστροφο πίνακα $\mathbf{W}^\#$, ο οποίος υπολογίζεται από την σχέση:

$$\mathbf{W}^\# = (\mathbf{W}^T \cdot \mathbf{W})^{-1} \cdot \mathbf{W}^T \quad (4.5)$$

Έτσι η αναδρομική σχέση προσδιορισμού των συντελεστών βαρύτητας του πίνακα διαχωρισμού $\mathbf{W}$ παίρνει την τελική μορφή:

$$\Delta \mathbf{W} \propto (\mathbf{W}^T)^\# - f(\mathbf{y}) \cdot \mathbf{x}^T \quad (4.6)$$

Η σχέση αυτή βασίζεται στην τεχνική της στοχαστικής βάθμωσης. Όσον αφορά στην επέκταση του αλγορίθμου (4.4) στην περίπτωση του υπέρ-ορισμένου διαχωρισμού αυτή μπορεί να υπολογιστεί με χρήση των δομών των ομάδων του Lie και χρησιμοποιώντας Riemannian metrics. Η πλήρης απόδειξη αυτών των αλγορίθμων μπορεί να βρεθεί στο [46] όπου αποδεικνύονται αναλυτικά οι σχέσεις επαναπροσδιορισμού ξεκινώντας από το κριτήριο της ελαχιστοποίησης της αμοιβαίας πληροφορίας, το οποίο ισοδυναμεί με ικανοποίηση του κριτηρίου της στατιστικής ανεξαρτησίας των σημάτων στην έξοδο του δικτύου διαχωρισμού.

Μια διαφορετική λύση στο πρόβλημα του υπέρ-ορισμένου διαχωρισμού σημάτων μπορεί να επιτευχθεί με την βοήθεια της Principal Component Analysis (PCA). Εφαρμόζεται σαν διαδικασία προεπεξεργασίας των σημάτων που λαμβάνονται από τους δέκτες ακολουθούμενη από την εφαρμογή κλασικών μεθόδων διαχωρισμού σημάτων που στηρίζονται στην Independent Component Analysis (ICA) [47]. Η τεχνική αυτή απομακρύνει την πληροφορία ετεροσυσχέτισης στα σήματα που προέρχεται από τους δέκτες με την PCA βαθμίδα και διατηρεί εκείνη την πληροφορία που είναι απαραίτητη για την λύση

του προβλήματος δηλαδή τις στοχαστικά ασυσχέτιστες συνιστώσες. Όταν δεν υπάρχει θόρυβος κατά την λήψη των σημάτων από τους M δέκτες, η PCA στο στάδιο προεπεξεργασίας μετασχηματίζει γραμμικώς τα M αρχικά σήματα σε N νέα σήματα αφού μόνο N κύριες συνιστώσες (*Principal Components*) είναι κυρίαρχες ενώ οι υπόλοιπες ($N-M$) είναι σχεδόν αμελητέες παρουσιάζοντας πολύ μικρή διασπορά. Τα N μετασχηματισμένα σήματα εισάγονται κατόπιν σε ένα δίκτυο διαχωρισμού για να προκύψουν τα N διαχωρισμένα σήματα στην έξοδο του που αντιστοιχούν στις εκτιμήσεις των αρχικών καθαρών σημάτων. Η τεχνική αυτή όπως έχει αποδειχτεί [47] μπορεί να δώσει πολύ καλά αποτελέσματα και στην περίπτωση που υπεισέρχεται θόρυβος κατά την καταγραφή των σημάτων από τους δέκτες.

Όπως αναφέρθηκε και στην εισαγωγή του κεφαλαίου, στις μετρήσεις μας για τον διαχωρισμό των σημάτων θα κάνουμε χρήση του αλγορίθμου fastICA για τους λόγους που αναφέρθηκαν στο Κεφάλαιο 2.

4.3 Ορισμός του PSNR

Η φράση **peak signal-to-noise ratio** ή **PSNR**, είναι ένας όρος για το λόγο ανάμεσα στη μέγιστη πιθανή ισχύ ενός σήματος και στην ισχύ του θορύβου που επηρεάζει την αξιοπιστία αυτής της αναπαράστασης. Επειδή πολλά σήματα έχουν ένα μεγάλο εύρος δυναμικού, το PSNR συχνά εκφράζεται σε κλίμακα decibel.

Το PSNR χρησιμοποιείται συχνά σαν ένα μέτρο της ποιότητας ανακατασκευής κωδικοποιητών συμπίεσης με απώλειες (π.χ. συμπίεση εικόνας). Το σήμα σε αυτή την περίπτωση είναι το αρχικό δεδομένο και ο θόρυβος είναι το σφάλμα που εισάγεται με την συμπίεση. Όταν συγκρίνουμε κωδικοποιητές συμπίεσης χρησιμοποιείται σαν μία προσέγγιση στην ανθρώπινη αντίληψη στην ποιότητα ανακατασκευής, έτσι σε ορισμένες περιπτώσεις μία ανακατασκευή μπορεί να είναι πιο κοντά στην αρχική σε σχέση με κάποια άλλη ακόμα κι αν έχει χαμηλότερο PSNR (ένα μεγαλύτερο PSNR λογικά θα υπονοούσε ότι η ανακατασκευή θα είναι υψηλότερης ποιότητας). Πρέπει να είμαστε ιδιαίτερα προσεκτικοί με το εύρος εγκυρότητας αυτού του μέτρου· είναι συμπερασματικά έγκυρο μόνο όταν χρησιμοποιείται για να συγκρίνουμε αποτελέσματα του ίδιου κωδικοποιητή και ίδιου περιεχομένου.

Εύκολα ορίζεται μέσω του μέσου τετραγωνικού σφάλματος (Mean squared error, MSE) το οποίο για δύο $m \times n$ μονόχρωμες εικόνες I και J , όπου η μία από τις δύο εικόνες θεωρείται σαν μία προσέγγιση θορύβου της άλλης ότι:

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i,j) - K(I,J)]^2$$

Το PSNR ορίζεται ως εξής:

$$\begin{aligned} PSNR &= 10 * \log_{10} \left(\frac{MAX_I^2}{MSE} \right) \\ &= 20 * \log_{10} \left(\frac{MAX_I}{\sqrt{MSE}} \right) \\ &= 20 * \log_{10}(MAX_I) - 10 * \log_{10}(MSE) \end{aligned}$$

4.4 Μετρήσεις-Αποτελέσματα

Στην περίπτωση του ανηχοϊκού θαλάμου, το πρόβλημα της μίξης των σημάτων ομιλίας παρουσιάζει την πιο απλή του μορφή, την περίπτωση της στιγμιαίας μίξης των σημάτων. Κατά τη διάρκεια των μετρήσεων χρησιμοποιήθηκαν διαδοχικά από δύο μέχρι πέντε μικρόφωνα.

Συγκεκριμένα ο πίνακας μίξης που χρησιμοποιήθηκε στο Matlab είναι ο εξής:

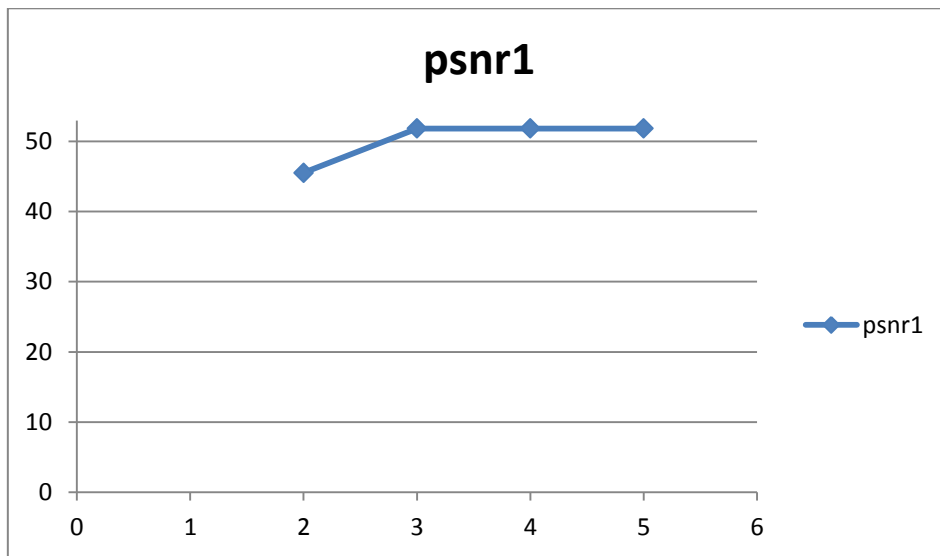
$$A=[0.5 \ 0.2; 0.3 \ 0.6; 0.4 \ 0.9; 0.5 \ 0.1; 0.3 \ 0.8]; \text{ \%πίνακας μίξης}$$

Δημιουργούμε στη συνέχεια τον πίνακα μίξης και κάνουμε διαχωρισμό σημάτων χρησιμοποιώντας τον αλγόριθμο fastICA. Μετά το διαχωρισμό των σημάτων υπολογίζουμε το PSNR ως εξής:

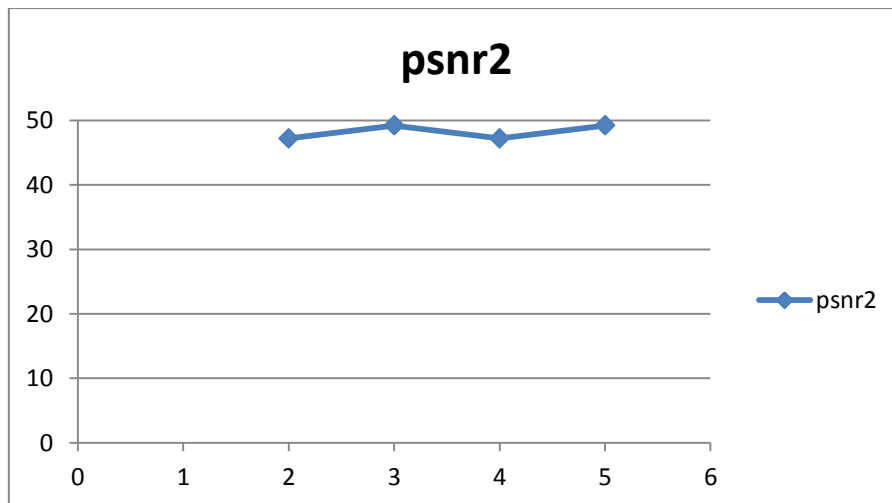
$$\begin{aligned} mse &= \text{mean2}((\text{single}(s) - \text{single}(c)).^2); \\ psnr &= 10 * \log_{10}((255^2)/mse); \end{aligned}$$

όπου «s» είναι το αρχικό μας σήμα και «c» το σήμα που προκύπτει μετά την εφαρμογή του fastICA.

Τα αποτελέσματα των μετρήσεων φαίνονται στα παρακάτω γραφήματα:



Σχήμα 4.2 Τιμές του PSNR για το πρώτο σήμα όταν έχουμε 2,3,4 ή 5 μικρόφωνα αντίστοιχα.



Σχήμα 4.3 Τιμές του PSNR για το δεύτερο σήμα όταν έχουμε 2,3,4 ή 5 μικρόφωνα αντίστοιχα.

Παρατηρούμε, πως η αύξηση του χρησιμοποιούμενου αριθμού μικροφώνων έχει ελάχιστη επίδραση στην ακρίβεια του συστήματος αναγνώρισης, αφού όπως διαπιστώνεται εύκολα από τις παραπάνω γραφικές παραστάσεις το ποσοστό αυτό είναι ήδη σε πολύ υψηλό επίπεδο ακόμα και μετά την χρήση μόνο δύο μικροφώνων.

Έτσι μπορούμε να συμπεράνουμε ότι η αύξηση του αριθμού των μικροφώνων στην περίπτωση του υπέρ-ορισμένου διαχωρισμού ταυτόχρονων ομιλητών, για την απλή περίπτωση της στιγμιαίας μίξης, δεν συμβάλλει σχεδόν καθόλου στην αύξηση του ποσοστού αναγνώρισης των διαχωρισμένων σημάτων ομιλίας καθώς παρόμοιας ακρίβειας αποτελέσματα μπορούν να

ληφθούν και με χρήση ίδιου αριθμού μικροφώνων με τους ταυτόχρονους ομιλητές.

Κεφάλαιο 5

Συνελικτική μίξη σημάτων με χρήση του μοντέλου PARAFAC

5.1 Εισαγωγή

Ο τυφλός διαχωρισμός σημάτων (BSS) έχει ως στόχο να υπολογίσει πολλαπλά σήματα πηγής που αναμιγνύονται μέσω ενός άγνωστου καναλιού, χρησιμοποιώντας μόνο τα παρατηρημένα σήματα που συλλαμβάνονται από ένα σύνολο αισθητήρων. Υπάρχουν διαφορετικές πιθανές εφαρμογές του τυφλού διαχωρισμού σημάτων σε διάφορες περιοχές, συμπεριλαμβανομένων την επεξεργασία ομιλίας, τις τηλεπικοινωνίες, τη βιοϊατρική επεξεργασία σήματος, την ανάλυση αστρονομικών στοιχείων ή δορυφορικών εικόνων, κ.λπ. Σε αυτό το κεφάλαιο, εστιάζουμε στον τυφλό διαχωρισμό σημάτων ομιλίας που καταγράφονται μέσα σε ένα περιβάλλον με ανακλάσεις. Σ' αυτή την περίπτωση, πολλαπλές εξασθενημένες αλλά και με καθυστέρηση εκδοχές του κάθε ομιλητή συλλαμβάνονται από κάθε μικρόφωνο, το οποίο καταλήγει σε ένα πρόβλημα τυφλού διαχωρισμού σημάτων συνελικτικής μίξης. Αυτό είναι πρόβλημα κλειδί σε εφαρμογές όπως η τηλεδιάσκεψη ή η κινητή τηλεφωνία, όπου ο πολλαπλός διαχωρισμός ομιλητών ή ο διαχωρισμός ομιλητή-υποβάθρου μπορεί να είναι σημαντικός για την ανθρώπινη σαφήνεια και την αυτόματη αναγνώριση ομιλίας.

Οι τεχνικές του τυφλού διαχωρισμού σημάτων συχνά θεωρούν συγκεκριμένες ιδιότητες στις πηγές ή στο σύστημα μίξης και βασίζονται σε ένα κριτήριο διαχωρισμού που επιβάλλει τις ίδιες ιδιότητες στους υπολογισμούς τους. Στον τυφλό διαχωρισμό σημάτων ομιλίας, μια σημαντική συνεισφορά που μπορεί να εκτιμηθεί είναι η υπάρχουσα μη στατικότητα αυτών των σημάτων. Τα σήματα ομιλίας πράγματι θεωρούνται μη στατικά για διάρκειες μεγαλύτερες των 40ms. Πολλοί αλγόριθμοι τυφλού διαχωρισμού σημάτων που εκμεταλλεύονται τη μη στατικότητα έχουν προταθεί στην απλή περίπτωση της στιγμιαίας γραμμικής μίξης. Στην πιο ρεαλιστική περίπτωση της γραμμικής συνελικτικής μίξης, στο πεδίο του χρόνου αλλά και στο πεδίο της συχνότητας έχουν προταθεί πολλές μέθοδοι. Αναφερόμαστε στο [48] για μια κατηγοριοποίηση των υπαρχόντων μεθόδων διαχωρισμού συνελικτικής μίξης [48].

Εκμεταλλευόμενοι τη μη στατική φύση των σημάτων ομιλίας, το πρόβλημα του τυφλού διαχωρισμού σημάτων μπορεί να λυθεί μέσω της χρήσης στατιστικών δεύτερης τάξης (SOS), υποθέτοντας ασυσχέτιστε πηγές. Έτσι το πρόβλημα περιορίζεται στον υπολογισμό του πίνακα μίξης που ελαχιστοποιεί ένα μέτρο της συνολικής ετεροσυσχέτισης. Αν το σύστημα μίξης είναι στατικό, η λύση μπορεί να αποκτηθεί αν θεωρήσουμε πολλαπλές καθυστερήσεις των ετεροσυσχετίσεων, κάτι που σημαίνει ότι πλέον έχουμε να αντιμετωπίσουμε ένα πρόβλημα από κοινού διαγωνιοποίησης (Joint-Approximate-Diagonalization-JAD). Μια τέτοια προσέγγιση προτάθηκε στο [49], για τυφλό διαχωρισμό σημάτων στιγμιαίας μίξης και στα [50], [51], [52], για τυφλό διαχωρισμό σημάτων συνελικτικής μίξης στο πεδίο της συχνότητας. Οι κύριες προκλήσεις προς τους αλγόριθμους εφαρμοσμένης μηχανικής για τον τυφλό διαχωρισμό σημάτων συνελικτικής μίξης στο πεδίο της συχνότητας είναι οι ακόλουθες.

- 1) Να δημιουργηθεί ένας γρήγορος και σθεναρός αλγόριθμος διαχωρισμού που λύνει το JAD πρόβλημα για κάθε συχνότητα.
- 2) Να εξεταστεί το υπο-ορισμένο πρόβλημα, δηλαδή, όταν ο αριθμός των πηγών είναι μεγαλύτερος από τον αριθμό των μικροφώνων. Αυτό συμπεριλαμβάνει θέματα αναγνώρισης και απαιτεί κατάλληλες τεχνικές μείωσης του cross-talk.
- 3) Να εξεταστούν αποτελεσματικά τα προβλήματα που αφορούν στην εξαρτημένη από τη συχνότητα ασάφεια λόγω αντιμετάθεσης και βάρθρωσης.
- 4) Να εξεταστούν τα μη στατικά περιβάλλοντα μίξης, δηλαδή να λυθεί το πρόβλημα του τυφλού διαχωρισμού σημάτων προσαρμόζοντάς το ανάλογα.

Σε αυτό το κεφάλαιο, προτείνουμε τις βασικές συνθήκες για κάθε μια από τις παραπάνω περιπτώσεις. Αρχικά, δείχνουμε ότι λύνοντας ένα JAD πρόβλημα για κάθε συχνότητα ισοδυναμεί με το να ταιριάξουμε ένα μοντέλο ενός συμμετρικά συζευγμένου παράλληλου παράγοντα (PARAFAC) με κάθε συχνότητα. Ο αλγόριθμος PARAFAC είναι ένα πανίσχυρο πολυγραμμικό εργαλείο άλγεβρας για διανυσματικό διαχωρισμό σε ένα άθροισμα πρώτης τάξεως διανυσμάτων. Μ' αυτήν την έννοια, ο αλγόριθμος PARAFAC είναι μια πιθανή γενίκευση του διαχωρισμού πρώτης τάξης (SVD) σε διανύσματα υψηλότερης τάξης. Ο PARAFAC παρουσιάστηκε το 1970 και σταδιακά άρχισε να χρησιμοποιείται σε διάφορους τομείς όπως των χημειομετρήσεων (Chemometrics) και της τεχνολογίας τροφίμων, της εξερευνητικής ανάλυσης δεδομένων, των ασύρματων επικοινωνιών και του τυφλού διαχωρισμού σημάτων. Στο κεφάλαιο αυτό, η χρήση της αλγεβρικής δομής του μοντέλου PARAFAC για κάθε συχνότητα επιτρέπει ένα βήμα μείωσης διάστασης πριν από το στάδιο του διαχωρισμού. Αυτό οδηγεί σε πολύ μικρότερη πολυπλοκότητα σε σχέση με τις JAD τεχνικές, με συγκεκριμένη σύγκλιση.

Στη συνέχεια, θα δείξουμε ότι, σε αντίθεση με τους κλασικούς αλγορίθμους από κοινού διαγωνιοποίησης, οι ισχυρές ιδιότητες μοναδικότητας του PARAFAC μας επιτρέπουν να προσδιορίσουμε τη συνάρτηση μεταφοράς του πίνακα μίξης σε συγκεκριμένες υπο-ορισμένες περιπτώσεις. Για την απλούστερη περίπτωση της στιγμιαίας μίξης, έχει καθιερωθεί ένα ανάλογο αποτέλεσμα [53]. Προτείνουμε την κατασκευή του πίνακα διαχωρισμού χρησιμοποιώντας την τεχνική του χρονικά μεταβαλλόμενου cross-talk με βάση τον σχεδιασμό Caron, και να αποδείξουμε την καλή επίδοση για τις υπο-ορισμένες περιπτώσεις.

Το τρίτο θέμα αυτού του κεφαλαίου είναι μια χαμηλής πολυπλοκότητας τεχνική για να αντιμετωπίσουμε το εξαρτημένο από τη συχνότητα πρόβλημα της αντιμετάθεσης. Η μέθοδος αυτή περιλαμβάνει την ομαδοποίηση των (ορθά κλιμακούμενων) εκτιμώμενων προφίλ πηγών μέσω του k -μέσου αλγορίθμου, μετά την οποία οι αντιμεταθετικοί πίνακες υπολογίζονται σε ένα μόνο στάδιο, με έναν μη-επαναληπτικό τρόπο. Η στρατηγική αυτή της ομαδοποίησης οδηγεί σε σημαντική μείωση της πολυπλοκότητας, σε σύγκριση με τις πλήρως επαναληπτικές τεχνικές που προτείνονται στο [52], [54], και [55], χωρίς να θυσιάζεται η απόδοση.

Τέλος, αποδεικνύουμε μια προσαρμοστική εκδοχή του αλγορίθμου μαζικής υλοποίησης του τυφλού διαχωρισμού, η οποία βασίζεται σε έναν από τους προσαρμοστικούς αλγορίθμους που έχουμε αναπτύξει στο [56] για να εντοπίζουμε τον διαχωρισμό PARAFAC. Αυτό είναι σημαντικό για τον εντοπισμό των αλλαγών στο ακουστικό περιβάλλον (π.χ., λόγω της μετακίνησης των ομιλητών), και θα αποφέρει επίσης μια μείωση στην πολυπλοκότητα, με αποτέλεσμα η συνολική θεωρητική λύση να είναι πιο κοντά στην πράξη.

Οι συμβολισμοί που χρησιμοποιούνται σε αυτό το κεφάλαιο είναι οι ακόλουθοι: Ένας τρίτης τάξης τανυστής μεγέθους $I \times J \times K$ συμβολίζεται με ένα καλλιγραφικό γράμμα $\mathcal{Y}$, και τα στοιχεία του και συμβολίζονται με $y_{i,j,k}, i = 1, \dots, I, j = 1, \dots, J$ και $k = 1, \dots, K$. Ένα κεφάλαιο γράμμα $\mathbf{Y}$ με έντονη γραφή υποδηλώνει έναν πίνακα και ένα πεζό γράμμα $\mathbf{y}$ με έντονη γραφή υποδηλώνει ένα διάνυσμα. Οι τροποποιημένοι, οι συζυγείς, οι συζυγείς τροποποιημένοι και οι ψευδο-αντίστροφοι σημειώνονται με $\mathbf{Y}^T, \mathbf{Y}^*, \mathbf{Y}^H$ και $\mathbf{Y}^\dagger$, αντίστοιχα. Το $\|\mathbf{Y}\|_F$ δηλώνει τη νόρμα Frobenius του $\mathbf{Y}$. Το αποτέλεσμα Kronecker συμβολίζεται με $\otimes$. Το αποτέλεσμα του Khatri-Rao συμβολίζεται με $\odot$, δηλ., $[a_1, \dots, a_I] \odot [b_1, \dots, b_I] = [a_1 \otimes b_1, \dots, a_I \otimes b_I]$. Ο $P \times P$ πίνακας ταυτοποίησης συμβολίζεται με $\mathbf{I}_P$. Το $E[\cdot]$ συμβολίζει τον διαχειριστή προσδοκίας. Θα χρησιμοποιήσουμε επίσης μια σημειογραφία τύπου Matlab για τους πίνακες υποσύνολα, δηλαδή, το $[\mathbf{A}]_{l:m,n:p}$ αντιπροσωπεύει τον πίνακα που δημιουργείται μετά την επιλογή των $m - l + 1$ σειρών του $\mathbf{A}$, από την l -οστή στην m -οστή, και οι $p - n + 1$ στήλες του $\mathbf{A}$, από την n -οστή στην p -οστή. Ο $[\mathbf{A}]_{:,n:p}$ χρησιμοποιείται για να υποδηλώσει την επιλογή όλων των γραμμών και ο $[\mathbf{A}]_{l:m,:}$ για να υποδηλώσει την επιλογή όλων των στηλών. Ομοίως, ο $\mathbf{y}(l:m)$ αντιπροσωπεύει μια επιλογή $m - l + 1$ δειγμάτων του διανύσματος $\mathbf{y}$, από το l -οστό στην m -οστό.

5.2 Διατύπωση του προβλήματος

5.2.1 Δεδομένα του προβλήματος

Έστω I σήματα ομιλίας $s(t) = [s_1(t), s_2(t), \dots, s_I(t)]$ από N ομιλητές που βρίσκονται σε ένα δωμάτιο και J μικρόφωνα τα οποία καταγράφουν τη συνελικτική τους μίξη $x(t) = [x_1(t), x_2(t), \dots, x_J(t)]$. Το χωρίς θόρυβο μοντέλο της συνελικτικής μίξης δίνεται από τη σχέση:

$$\mathbf{x}(t) = \mathbf{H} * \mathbf{s}(t) = \sum_{l=0}^{L-1} \mathbf{H}(l) \mathbf{s}(t-l) \quad (5.1)$$

όπου $*$ είναι ο τελεστής της γραμμικής συνέλιξης. Ο $J \times I$ πίνακας αναπαριστά το σύστημα μίξης με χρονική καθυστέρηση l . Τα στοιχεία του $h_{j,i}(l)$ είναι συντελεστές που προκύπτουν λόγω των ανακλάσεων μέσα στο δωμάτιο (RIR-Room Impulse Response) μεταξύ της πηγής i και του μικροφώνου j , με βάση το

μοντέλο ενός FIR(Finite Impulse Response) φίλτρου. Το L είναι το μέγιστο (άγνωστο) μήκος του καναλιού. Για να υπολογίσουμε τις πηγές $\hat{s}(t) = [\hat{s}_1(t), \hat{s}_2(t), \dots, \hat{s}_I(t)]^T$, στόχος είναι να βρούμε έναν πίνακα αντίστροφου-καναλιού $\mathbf{W}$ με διαστάσεις $I \times J$, τέτοιος ώστε:

$$\hat{s}(t) = \mathbf{W} * \mathbf{x}(t) = \sum_{k=0}^{K-1} \mathbf{W}(k) \mathbf{x}(t - k) \quad (5.2)$$

όπου k είναι το μήκος του αντίστροφου καναλιού της κρουστικής απόκρισης.

Για να λυθεί αυτό το πρόβλημα, μπορούμε να καταφύγουμε σε μια προσέγγιση στο πεδίο του χρόνου ή σε μια προσέγγιση στο πεδίο της συχνότητας. Στο πεδίο του χρόνου, το k θα πρέπει να είναι τουλάχιστον ίσο με το μήκος του πραγματικού καναλιού L για να μοντελοποιηθούν όλες οι ανακλάσεις, και πολύ μεγαλύτερο από το L για τον ακριβή υπολογισμό. Οι μέθοδοι που χρησιμοποιούνται στο πεδίο του χρόνου είναι ευαίσθητοι στην αναντιστοιχία της τάξης των καναλιών, και οι ιδιότητες ταυτοποίησης δεν είναι επαρκώς κατανοητές, ειδικά για τις υπο-ορισμένες περιπτώσεις.

Οι μέθοδοι του τυφλού διαχωρισμού σημάτων στο πεδίο τη συχνότητας ξεκινούν με τη χαρτογράφηση του προβλήματος στο πεδίο της συχνότητας εφαρμόζοντας τον διακριτό μετασχηματισμό Fourier (discrete-Fourier transform-DFT) στα παρατηρηθέντα σήματα

$$\mathbf{x}(t) = \mathbf{H} * \mathbf{s}(t) \leftrightarrow \mathbf{x}(f, q) \approx \mathbf{H}(f) \mathbf{s}(f, q) \quad (5.3)$$

όπου f είναι ο δείκτης της συχνότητας, $f = 1, \dots, F$, το q είναι ο δείκτης των πλαισίων, το $\mathbf{x}(f, q) = [\mathbf{x}_1(f, q), \dots, \mathbf{x}_J(f, q)]^T$, και το $\mathbf{s}(f, q) = [\mathbf{s}_1(f, q), \dots, \mathbf{s}_I(f, q)]^T$. Η i -οστή στήλη του $\mathbf{H}(f)$ αντιπροσωπεύει την χωρική υπογραφή του i -οστού ομιλητή στο πεδίο της συχνότητας, σε συχνότητα f . Να σημειωθεί ότι η προσέγγιση που γίνεται στην εξίσωση 5.3 είναι ακριβής μόνο για περιοδικά σήματα $\mathbf{s}(t)$, ή ισοδύναμα, αν η χρονική συνέλιξη είναι κυκλική. Η προσέγγιση είναι ικανοποιητική αν το F είναι πολύ μεγαλύτερο από το μέγιστο μήκος L των αναμειγμένων καναλιών. Για να περιορίσουμε την επίδραση της κυκλικότητας, χρησιμοποιούμε μια φασματική προσέγγιση. Στην πράξη, θα υπολογίσουμε τον διακριτό μετασχηματισμό Fourier διαδοχικών επικαλυπτόμενων παραθυροποιημένων πλαισίων (συγκεκριμένα θα χρησιμοποιήσουμε ένα παράθυρο Hanning).

Το κύριο πλεονέκτημα της προσέγγισης στο πεδίο της συχνότητας είναι η μετατροπή της αρχικής συνέλιξης στο πεδίο του χρόνου σε ένα σύνολο στιγμιαίων προβλημάτων τυφλού διαχωρισμού σημάτων, για την επίλυση των οποίων έχουν προταθεί διάφοροι αλγόριθμοι. Ωστόσο, η κυριότερη δυσκολία με τον τυφλό διαχωρισμό σημάτων στο πεδίο της συχνότητας είναι η ανάγκη να αντιμετωπίσουμε τις ασάφειες λόγω αντιμετάθεσης και βάθμωσης, δηλαδή ο πίνακας μίξης θεωρείται με βάση μια αυθαίρετη αντιμετάθεση και βάθμωση των στηλών για κάθε συχνότητα. Πριν μετατρέψουμε ξανά τα κατ' εκτίμηση σήματα πηγής στο πεδίο του χρόνου, πρέπει να αντισταθμιστεί η ασάφεια της βάθμωσης και να εφαρμοστεί μια διαδικασία αντιπαραβολής της αντιμετάθεσης έτσι ώστε να συνδεθεί κάθε φασματικός παράγοντας με την αντίστοιχη πηγή.

Πολλές μέθοδοι έχουν προταθεί για την αντιμετώπιση της αντιμεταθετικής ασάφειας. Παρακάτω προτείνουμε μια νέα παραλλαγή των τεχνικών που έχουν ήδη προταθεί για τη διόρθωση της αντιμεταθετικής ασάφειας [52], [54], [55]. Αυτό έχει ως αποτέλεσμα τη σημαντική μείωση της πολυπλοκότητας που σχετίζεται με τις πλήρως επαναληπτικές μεθόδους [52], [54], [55], χωρίς να μειώνεται η απόδοση.

Πριν προχωρήσουμε περεταίρω, θέτουμε τις βασικές μας υποθέσεις.

Υπόθεση 1: Τα σήματα ομιλίας $s(t)$ έχουν μηδενική μέση τιμή και είναι αμοιβαία ασυσχέτιστα.

Υπόθεση 2: Ο αριθμός των ομιλητών I είναι γνωστός, αλλά όχι απαραίτητα μικρότερος από τον αριθμό των μικροφώνων J .

Υπόθεση 3: Οι κρουστικές αποκρίσεις όλων των φίλτρων μίξης θεωρούνται συνεχείς κατά τη διάρκεια των καταγραφών.

5.2.2 Υπολογισμός του καναλιού

Θεωρούμε ότι κάθε καταγεγραμμένο σήμα $x_j(t), j = 1, \dots, J$ είναι ένα διάνυσμα N δειγμάτων. Ας διαιρέσουμε όλο το σύνολο δεδομένων σε P μη επικαλυπτόμενα υποσύνολα, τέτοια ώστε κάθε υποσύνολο να περιέχει $N_p = \lfloor N/P \rfloor$ στιγμιότυπα. Αυτά τα υποσύνολα συμβολίζονται με $p = 1, \dots, P$, και το p -οστό υποσύνολο αντιστοιχεί σε ένα σύνολο N_p στιγμιότυπων μεταξύ των χρονικών στιγμών t_{p-1} και t_p . Συμβολίζουμε με $T_p = (N_p/F_s)$ τη διάρκεια κάθε υποσυνόλου, όπου F_s είναι η συχνότητα του δείγματος. Συνεπώς ο πίνακας αυτοσυσχέτισης $\mathbf{R}_x(f, p) \stackrel{\text{def}}{=} E[\mathbf{x}(f, p)\mathbf{x}^H(f, p)]$ τάξης $J \times J$ μπορεί να γραφεί ως εξής:

$$\mathbf{R}_x(f, p) = \mathbf{H}(f)\mathbf{R}_s(f, p)\mathbf{H}^H(f) \quad (5.4)$$

όπου $\mathbf{R}_s(f, p) \stackrel{\text{def}}{=} E[\mathbf{s}(f, p)\mathbf{s}^H(f, p)]$ είναι ο πίνακας αυτοσυσχέτισης των σημάτων ομιλίας στο p -οστό υποσύνολο για συχνότητα f . Αλγόριθμοι που εκμεταλλεύονται τη μη στατικότητα πρέπει να επιλέξουν T_p τέτοιο ώστε τα διαδοχικά υποσύνολα να είναι ασυσχέτιστα. Για εφαρμογές σε σήματα ομιλίας, η διάρκεια του υποσυνόλου T_p θα πρέπει να είναι τουλάχιστον 40ms, αφού αυτός είναι ο μικρότερος χρόνος για τον οποίο τα σήματα θεωρούνται μη στατικά[1]. Οι στατιστικές είναι τότε αρκετά διαφορετικές από τη μια καθυστέρηση στην άλλη, τέτοιες ώστε η κάθε μια μπορεί ταυτόχρονα να εκμεταλλευτεί τα P υποσύνολα, για μια δεδομένη συχνότητα.

$$\begin{cases} \mathbf{R}_x(f, 1) = \mathbf{H}(f)\mathbf{R}_s(f, 1)\mathbf{H}^H(f) \\ \vdots \\ \mathbf{R}_x(f, P) = \mathbf{H}(f)\mathbf{R}_s(f, P)\mathbf{H}^H(f) \end{cases} \quad (5.5)$$

Εφόσον υποθέτουμε αμοιβαία ασυσχέτιστα σήματα ομιλίας, θεωρούμε διαγώνιους πίνακες αυτοσυσχέτισης $\mathbf{R}_s(f, p)$, για $f = 1, \dots, F$ και $p = 1, \dots, P$. Έτσι ο υπολογισμός του $\mathbf{H}(f)$ διαμορφώνεται σαν ένα JAD πρόβλημα, για κάθε συχνότητα.

Στην πράξη, οι ακριβείς πίνακες αυτοσυσχέτισης $\mathbf{R}_x(f, P)$ δεν είναι διαθέσιμοι, μπορούν όμως να υπολογιστούν από τα δείγματα του $\mathbf{x}(t), t = 1, \dots, N$. Για κάθε υποσύνολο p των N_p δειγμάτων, υπολογίζουμε το σημείο F του διακριτού μετασχηματισμού Fourier των διαδοχικών επικαλυπτόμενων πλαισίων $\mathbf{x}(f, q)$ (το κάθε ένα αποτελείται από F χρονικά δείγματα) με ένα παράθυρο σημείου F (παράθυρο Hanning). Για παράδειγμα, αν συμβολίσουμε με α τον παράγοντα επικάλυψης (π.χ. $\alpha=0,75$), τότε ο αριθμός των επικαλυπτόμενων πλαισίων μέσα σε κάθε υποσύνολο p θα είναι:

$$M = \left\lfloor \frac{N_p - N_a}{F - N_a} \right\rfloor \quad (5.6)$$

όπου $[N_a = \alpha F]$ είναι ο αριθμός των δειγμάτων στο επικαλυπτόμενο τεμάχιο. Η εκτίμηση της μήτρας αυτοσυσχέτισης δειγμάτων, για τη συχνότητα f και το υποσύνολο p , δίνεται ως εξής:

$$\hat{\mathbf{R}}_x(f, P) = \frac{1}{M} \sum_{m=1}^M \mathbf{x}(f, k_{p,m} + 1 : k_{p,m} + F) \times \mathbf{x}^H(f, k_{p,m} + 1 : k_{p,m} + F) \quad (5.7)$$

όπου $k_{p,m}$ είναι ένας δείκτης που συνδυάζει τα p και m ως εξής:

$$k_{p,m} = (p - 1)N_p + (m - 1)(F - N_a) \quad (5.8)$$

Οι τεχνικές που στηρίζονται στην από κοινού διαγωνιοποίηση (JAD) και απαιτούν βαθμό $(\mathbf{H}(f))=I$, για $f = 1, \dots, F$, συνεπώς δεν μπορούν να χρησιμοποιηθούν στην περίπτωση του υπο-ορισμένου προβλήματος, όταν δηλαδή $J < I$. Στην επόμενη ενότητα, θα δείξουμε ότι κάθε σύστημα από κοινού διαγωνιοποίησης (εξίσωση 5.5) μπορεί ισοδύναμα να γραφεί σαν τον PARAFAC διαχωρισμό του διανύσματος $\mathbf{R}_x(f) \in \mathbb{C}^{J \times J \times P}$ που είναι τρίτης τάξης, και το οποίο κατασκευάζεται αν κατά μήκος της τρίτης διάστασης ενώσουμε τους πίνακες $P \{ \mathbf{R}_x(f, 1), \dots, \mathbf{R}_x(f, P) \}$ μεταξύ τους. Αυτή η αναδιατύπωση που βασίζεται στο μοντέλο PARAFAC χρησιμοποιήθηκε στο [53] για την περίπτωση της στιγμιαίας μίξης. Η γενίκευσή του για την περίπτωση της συνελκτικής μίξης υπονοεί ότι το μοντέλο PARAFAC είναι πλέον έγκυρο για κάθε συχνότητα. Ένα σημαντικό πλεονέκτημα της αναδιατύπωσης αυτού του μοντέλου πέρα από τις προαναφερθείσες τεχνικές της από κοινού διαγωνιοποίησης είναι ότι δεν απαιτεί το $J > I$ για να είναι ο πίνακας μίξης μοναδικός.

5.3 Σύνδεση με το μοντέλο PARAFAC

5.3.1 Αναδιατύπωση του προβλήματος

Σε αυτή την ενότητα δείχνουμε ότι η εξίσωση 5 ισοδυναμεί με το μοντέλο PARAFAC. Κάθε στοιχείο του διανύσματος $\mathbf{R}_x(f)$ συμβολίζεται με $r_{j_1, j_2, p}^x(f)$ με $j_1 = 1, \dots, J$, $j_2 = 1, \dots, J$ και $p = 1, \dots, P$. Τα στοιχεία του $\mathbf{H}(f)$ συμβολίζονται με $h_{j,i}(f)$. Κατασκευάζουμε τον $P \times I$ πίνακα $\mathbf{C}(f)$, και το

στοιχείο του που βρίσκεται στην p -οστή σειρά και στην i -οστή στήλη, και συμβολίζεται με $c_{p,i}(f)$, είναι το i -οστό διαγώνιο στοιχείο του $\mathbf{R}_S(f, p)$, δηλαδή η φασματική πυκνότητα ισχύος της i -οστής πηγής μεταξύ του p -οστού υποσυνόλου με συχνότητα f . Συνεπώς τα στοιχεία $r_{j_1, j_2, p}^x(f)$ μπορούν να γραφούν σαν ένα άθροισμα όπως φαίνεται παρακάτω:

$$r_{j_1, j_2, p}^x(f) = \sum_{i=1}^I h_{j_1, i}(f) c_{p, i}(f) h_{j_2, i}^*(f) \quad (5.9)$$

Η εξίσωση 9 είναι γνωστή ως συμμετρικά συζευγμένος διαχωρισμός PARAFAC του διανύσματος και ο αριθμός των στοιχείων είναι ο βαθμός αυτού του διανύσματος [57]. Υπολογίζοντας τον διαχωρισμό PARAFAC του $\mathbf{R}_x(f)$ ανεξάρτητα για κάθε συχνότητα, αποκτούμε ολόκληρη τη συλλογή των πινάκων μίξης στο πεδίο του χρόνου $\{\mathbf{H}(f), f = 1, \dots, F\}$ και το φάσμα ισχύος των πηγών $\{\mathbf{C}(f), f = 1, \dots, F\}$, μέχρι την αντιμετάθεση που εξαρτάται από τη συχνότητα και την βάρθρωση των στηλών. Παρακάτω, θα περιγράψουμε τις μοναδικές συνθήκες για τον συμμετρικά συζευγμένο PARAFAC, με βάση τις οποίες αυτές οι μήτρες είναι αναγνωρίσιμες μέχρι να συναντήσουμε τις υπάρχουσες απροσδιοριστίες.

5.3.2 Αναγνωρισιμότητα

Ο τανυστής $\mathbf{R}_x(f)$ κατασκευάζεται από το συνδυασμό των στοιχείων των πινάκων $\mathbf{H}(f)$ και $\mathbf{C}(f)$ (εξίσωση 9). Ο συμμετρικά συζευγμένος διαχωρισμός PARAFAC του $\mathbf{R}_x(f)$ στην εξίσωση 9 θεωρείται ότι είναι μοναδικός αν κάθε ζευγάρι των πινάκων $\tilde{\mathbf{H}}(f)$ και $\tilde{\mathbf{C}}(f)$ που ικανοποιεί την εξίσωση 9 σχετίζεται με τους $\mathbf{H}(f)$ και $\mathbf{C}(f)$ μέσω της σχέσης:

$$\mathbf{H}(f) = \tilde{\mathbf{H}}(f) \mathbf{\Pi} \mathbf{\Lambda}_1, \quad \mathbf{C}(f) = \tilde{\mathbf{C}}(f) \mathbf{\Pi} \mathbf{\Lambda}_2 \quad (5.10)$$

με $\mathbf{\Lambda}_1, \mathbf{\Lambda}_2$ διαγώνιοι πίνακες που ικανοποιούν την $\mathbf{\Lambda}_1 \mathbf{\Lambda}_1^* \mathbf{\Lambda}_2 = \mathbf{I}_1$ και $\mathbf{\Pi}$ ένας αντιμεταθετικός πίνακας. Ως εκ τούτου, οι ασάφειες του μοντέλου PARAFAC είναι οι ίδιες με αυτές του προβλήματος της από κοινού διαγωνιοποίησης (JAD), δηλαδή, οι $\mathbf{H}(f)$ και $\mathbf{C}(f)$ υπολογίζονται με αυθαίρετη βάρθρωση και αντιμετάθεση των στηλών τους. Ο τρόπος με τον οποίο αυτές οι ασάφειες μπορούν να διορθωθούν θα συζητηθεί στην ενότητα 4.

Ένα πρώτο αποτέλεσμα μοναδικότητας απαιτεί το θεώρημα του βαθμού-Kruskal ενός πίνακα [57].

1^{ος} ορισμός: Ο βαθμός Kruskal ή k -βαθμός ενός πίνακα $\mathbf{H}$, που συμβολίζεται με $k_{\mathbf{H}}$, είναι ο μέγιστος αριθμός r τέτοιος ώστε κάθε σύνολο r στηλών του $\mathbf{H}$ να σχηματίζει ένα γραμμικά ανεξάρτητο σύνολο.

Το ακόλουθο θεώρημα καθορίζει μια κατάσταση κάτω από την οποία εγγυάται η μοναδικότητα του συμμετρικά συζευγμένου διαχωρισμού PARAFAC [57],[58].

1^ο Θεώρημα: Ο διαχωρισμός (εξίσωση 9) είναι μοναδικός αν

$$2k_{H(f)} + k_{C(f)} \geq 2(I + 1) \quad (5.11)$$

Αξίζει να σημειωθεί η παραπάνω συνθήκη είναι ικανή αλλά όχι αναγκαία για την αναγνωρισιμότητα. Για μια διαφορετική συνθήκη μοναδικότητας υποθέτουμε ότι $I < P$. Μια άλλη συνθήκη για τη μοναδικότητα του συμμετρικά συζευγμένου διαχωρισμού PARAFAC φαίνεται με το ακόλουθο θεώρημα [59]:

2^ο *Θεώρημα*: Υποθέτουμε ότι έχουμε πάρει τα στοιχεία των $H(f)$ και $C(f)$ από μια από κοινού συνεχή κατανομή. Αν $I \leq P$ και

$$\frac{I(I-1)}{2} \leq \frac{J(J-1)}{4} \left(\frac{J(J-1)}{2} + 1 \right) - \frac{J!}{(J-4)! 4!} 1_{\{J \geq 4\}} \quad (5.12)$$

όπου

$$1_{\{J \geq 4\}} = \begin{cases} 0, & \text{αν } J < 4 \\ 1, & \text{αν } J \geq 4 \end{cases}$$

Τότε οι $H(f)$ και $C(f)$ είναι μοναδικοί με πιθανότητα ένα.

Στην παρούσα εργασία, το J αντιπροσωπεύει τον αριθμό των μικροφώνων και το I τον αριθμό των πηγών. Ο ακόλουθος πίνακας δείχνει το άνω όριο για το I τέτοιο ώστε να ικανοποιείται η εξίσωση 12, για διαφορετικές τιμές του J [53].

J	2	3	4	5	6	7	8
I_{MAX}	2	4	6	10	15	20	26

Από αυτόν τον πίνακα, είναι εμφανές ότι η PARAFAC αναδιατύπωση του τυφλού διαχωρισμού σημάτων στο πεδίο της συχνότητας επιτρέπει, θεωρητικά, μοναδική αναγνώριση των $H(f)$ πινάκων μίξης, για $f = 1, \dots, F$, ακόμα και σε υπο-ορισμένες περιπτώσεις. Αυτό είναι ένα σημαντικό πλεονέκτημα μπροστά στις τεχνικές διαγωνιοποίησης, οι οποίες για να λυθούν απαιτούν $J \geq I$ (εξίσωση 5.5). Να σημειώσουμε επίσης ότι με το να επικαλούμαστε τις μοναδικές ιδιότητες του PARAFAC είναι ένας τρόπος να αποδείξουμε με σαφήνεια ότι η από κοινού αποσυσχέτιση ενός συνόλου πινάκων είναι ένα επαρκές κριτήριο για τον μοναδικό διαχωρισμό.

Στην επόμενη ενότητα, περιγράφουμε τη μαζική υλοποίηση (Batch Implementation) του PARAFAC διαχωρισμού για να διαχωρίσουμε τις πηγές στο πεδίο της συχνότητας, σε ένα στατικό περιβάλλον μίξης.

5.4 Μαζική υλοποίηση (Batch PARAFAC)

5.4.1 Απεικόνιση του διανύσματος σε πίνακα

Πολλοί από τους αλγορίθμους που έχουν σχεδιαστεί για να υπολογίσουν τον PARAFAC διαχωρισμό ενός διανύσματος χρησιμοποιούν τον διαφορετικό πίνακα αναπαράστασης αυτού του διανύσματος. Στην παρούσα εργασία, θα χρησιμοποιήσουμε την ακόλουθη αναπαράσταση του $R_x(f)$ που είναι ένας πίνακας μεγέθους $J^2 \times P$:

$$[\mathbf{R}_x(f)]_{(j_1-1)J+j_2,p} = [\mathbf{R}_x(f)]_{j_1,j_2,p} \quad (5.13)$$

με $j_1 = 1, \dots, J$, $j_2 = 1, \dots, J$ και $p = 1, \dots, P$. Λόγω του συμμετρικά συζευγμένου μοντέλου PARAFAC, το $\mathbf{R}_x(f)$ συνδέεται με τους άγνωστους πίνακες $\mathbf{H}(f)$ και $\mathbf{C}(f)$ ως εξής:

$$\mathbf{R}_x(f) = [\mathbf{H}(f) \odot \mathbf{H}^*(f)] \mathbf{C}^T(f) \quad (5.14)$$

5.4.2 Υπολογισμός του διαχωρισμού PARAFAC

Για να υπολογίσουμε τους πίνακες $\mathbf{H}(f)$ και $\mathbf{C}(f)$ που ταιριάζουν κατάλληλα στο $\mathbf{R}_x(f)$ του μοντέλου PARAFAC, χρησιμοποιούμε έναν εναλλάσων αλγόριθμο ελαχίστων τετραγώνων (ALS). Η ιδέα του ALS είναι να αναβαθμίζει αυτούς τους πίνακες με έναν εναλλασσόμενο τρόπο σε κάθε επανάληψη. Μπορούμε προσωρινά να αγνοήσουμε τη συμμετρία στο μοντέλο, δηλαδή να χρησιμοποιήσουμε τον $\mathbf{H}(f)$ και τον $\mathbf{H}^H(f)$ σαν ανεξάρτητες μεταβλητές. Η συζευγμένη συμμετρία των δεδομένων στην εξίσωση 5.14 διασφαλίζει ότι υπάρχει μικρό σφάλμα εφόσον το κάνουμε αυτό· στο τέλος μπορούμε να χρησιμοποιήσουμε τον ένα από τους δυο πίνακες για να υπολογίσουμε τον $\mathbf{H}(f)$ ή να υπολογίσουμε τον μέσο όρο τους. Περισσότερες πληροφορίες για τον ALS υπάρχουν στο [60], [61] και [62]. Το πλεονέκτημα του ALS είναι ότι δουλεύει για τις κατώτατες συνθήκες· έχει όμως αργή σύγκλιση όταν έχουμε ανεπαρκώς ρυθμισμένα δεδομένα. Ένα ενισχυμένο σχέδιο αναζήτησης μπορεί να εισαχθεί στο βρόχο ALS για να επιταχύνει τη σύγκλιση, όπως προτείνεται στο [63] για πραγματικές συνθήκες και στο [64] για πιο πολύπλοκες περιπτώσεις. Επίσης, μπορεί κανείς να καταφύγει σε τεχνικές βελτιστοποίησης όπως ο αλγόριθμος Levenberg-Marquardt [65]. Αξίζει να σημειωθεί ότι η πολυπλοκότητα αυτών των αλγορίθμων μπορεί να μειωθεί σημαντικά με ένα βήμα προεπεξεργασίας που μειώνει τη διάσταση [34]. Ένας άλλος πολύ αποτελεσματικός αλγόριθμος για τον υπολογισμό του PARAFAC διαχωρισμού προτάθηκε στο [66] και χρησιμοποιήθηκε στα [53],[67]. Ο αλγόριθμος αυτός, που θα τον ονομάσουμε PARAFAC μέσω ταυτόχρονης διαγωνιοποίησης (PARAFAC via Simultaneous Diagonalization) ή PARAFAC-SD υπολογίζει τον PARAFAC διαχωρισμό ενός διανύσματος $\mathbf{R} \in \mathbb{C}^{J_1 \times J_2 \times J_3}$ I – βαθμού μέσω της από κοινού διαγωνιοποίησης ενός συνόλου I συμμετρικών πινάκων, μεγέθους $I \times I$. Μπορεί να εφαρμοστεί μόνο στην περίπτωση που $I \leq \min(J_1, J_2, J_3)$, όπου ο ρόλος των J_1, J_2 και J_3 μπορεί να αλλάξει. Αυτή η συνθήκη χρησιμοποιείται συχνά στην πράξη, όπου ο χρόνος είναι τυπικά η μεγαλύτερη διάσταση J_3 του παρατηρούμενου διανύσματος. Λόγω της υψηλής ακρίβειας και της χαμηλής πολυπλοκότητας, ο αλγόριθμος PARAFAC-SD είναι ένας καλός σύμμαχος για την επίλυση του προβλήματος του τυφλού διαχωρισμού σημάτων σε αυτή την εργασία. Ας περιγράψουμε εν συντομία την αρχή αυτού του αλγορίθμου. Υποθέτουμε ότι $I \leq \min(J^2, P)$, το οποίο αποτελεί μια ρεαλιστική υπόθεση για το πρόβλημα του τυφλού διαχωρισμού. Ας θεωρήσουμε τον πίνακα $\mathbf{R}_x(f) \in \mathbb{C}^{J_1 \times J_2 \times J_3}$ της εξίσωσης 5.14. Αν

βαθμός($\mathbf{H}(f)$) = $\min(I, J)$, τότε λόγω των αποτελεσμάτων των Hharti-Rao, θα έχουμε βαθμός($\mathbf{H}(f) \odot \mathbf{H}^*(f)$) = I . Υποθέτοντας ότι $P \geq I$, το $\mathbf{C}(f)$ είναι γενικά I -βαθμού. Συνεπώς, το $\mathbf{R}_x(f)$ είναι I -βαθμού και η μειωμένου μεγέθους μοναδική τιμή διαχωρισμού (SVD) μπορεί να γραφεί ως εξής:

$$\mathbf{R}_x(f) = \mathbf{U}(f)\mathbf{\Sigma}(f)\mathbf{V}^H(f) \quad (5.15)$$

όπου $\mathbf{U}(f) \in \mathbb{C}^{J^2 \times I}$, $\mathbf{\Sigma}(f) \in \mathbb{R}^{I \times I}$ είναι διαγώνιος και $\mathbf{V}(f) \in \mathbb{C}^{P \times I}$. Ας σημειωθεί ότι όταν ο αριθμός των ομιλητών I είναι *a priori* άγνωστος, μπορεί να εκτιμηθεί ως ο αριθμός των μοναδικών τιμών του $\mathbf{R}_x(f)$, για μια δοσμένη f . Η βασική ιδέα του PARAFAC-SD είναι να συνδέσουμε τις εξισώσεις 14 και 15. Θεωρώντας ότι το $\mathbf{R}_x(f)$ είναι I -βαθμού, υπάρχει τότε ένας μη-μοναδικός πίνακας $\mathbf{Z}(f) \in \mathbb{R}^{I \times I}$, τέτοιος ώστε

$$\begin{cases} \mathbf{H}(f) \odot \mathbf{H}^*(f) = \mathbf{U}(f)\mathbf{\Sigma}(f)\mathbf{Z}(f) \\ \mathbf{C}^T(f) = \mathbf{Z}^{-1}(f)\mathbf{V}^H(f) \end{cases} \quad (5.16)$$

Ο υπολογισμός του $\mathbf{Z}(f)$ είναι επαρκής για να υπολογίσουμε τον PARAFAC διαχωρισμό. Προφανώς, $\mathbf{C}(f) = \mathbf{V}^*(f)\mathbf{Z}^{-T}(f)$. Επιπλέον, οι στήλες του $\mathbf{H}(f) \odot \mathbf{H}^*(f)$ είναι τα διανύσματα $\mathbf{h}_i(f) \otimes \mathbf{h}_i^*(f)$, όπου $i = 1, \dots, I$, τα οποία είναι οι διανυσματικές αναπαραστάσεις των πρώτου βαθμού $\mathbf{h}_i(f) \otimes \mathbf{h}_i^H(f)$ πινάκων. Συνεπώς, το $\mathbf{h}_i(f), i = 1, \dots, I$, μπορεί να οριστεί, μέχρι ενός κλιμακούμενο παράγοντα, ως το αριστερό μοναδικό διάνυσμα που σχετίζεται με τη μεγαλύτερη μοναδική τιμή του αντίστοιχου μοναδικού πίνακα, πρώτου βαθμού. Το σημείο κλειδί για να υπολογίσουμε τον $\mathbf{Z}(f)$ είναι να θεωρήσουμε ότι το $\mathbf{U}(f)\mathbf{\Sigma}(f)\mathbf{Z}(f)$ έχει μια Khattri-Rao δομή. Έχει αποδειχθεί [34], για τον μη συμμετρικό διαχωρισμό PARAFAC ότι με μαθηματική αναλογία το $\mathbf{Z}(f)$ διαγωνιοποιεί ένα σύνολο συμμετρικών $I \times I$ πινάκων $\{\mathbf{M}_1(f), \dots, \mathbf{M}_I(f)\}$. Για περεταίρω λεπτομέρειες για τον τρόπο με τον οποίο κατασκευάζονται αυτοί οι πίνακες, παραπέμπουμε στα [53], [66] και [67].

Αυτή η αναδιατύπωση έχει δύο σημαντικά πλεονεκτήματα επί των κλασικών αλγορίθμων τυφλού διαχωρισμού σημάτων που στηρίζονται στην από κοινού διαγωνιοποίηση. 1) ο PARAFAC είναι μοναδικά αναγνωρίσιμος σε συγκεκριμένες υπο-ορισμένες καταστάσεις (δες 5.3.2), *αποδεικνύοντας* έτσι τη μοναδικότητα του (εκτιμώμενου) πίνακα καναλιού, 2) ενώ οι JAD τεχνικές διαγωνιοποιούν το αρχικό σύστημα των P πινάκων μεγέθους $J \times J$, ο PARAFAC-SD χρησιμοποιεί πλήρως την ισχυρή αλγεβρική δομή του PARAFAC μοντέλου για να καταλήξει σε ένα μικρότερο JAD πρόβλημα συμπεριλαμβάνοντας I πίνακες, μεγέθους $I \times I$. Η μείωση της πολυπλοκότητας που προκύπτει είναι πολύ σημαντική, ακόμα και για μικρά σήματα. Ας θεωρήσουμε ένα απλό παράδειγμα με $J = 4$ μικρόφωνα και $I = 2$ ομιλητές, και ένα μικρό σπλιτ σήματος $P = 12$ epochs. Για κάθε συχνότητα, αντί να διαγωνιοποιήσουμε 12 πίνακες μεγέθους 4×4 ο καθένας, ο PARAFAC-SD διαγωνιοποιεί 2 πίνακες μεγέθους 2×2 . Με ένα μεγάλο μήκος FFT (π.χ. 1024), το πλεονέκτημα επί των κλασικών μεθόδων είναι σημαντικό.

Το πρόβλημα για κάθε συχνότητα μπορεί να λυθεί από κάθε αλγόριθμο JAD (ή PARAFAC). Η συνολική ακρίβεια του PARAFAC-SD εξαρτάται από τον αλγόριθμο που χρησιμοποιείται γι' αυτό το τελευταίο βήμα. Στην πράξη,

θα χρησιμοποιήσουμε την εκτεταμένη QZ-επανάληψη [69], όπως παρουσιάζεται στο [66].

Μόλις τελειώσει ο διαχωρισμός που βασίζεται στον PARAFAC, θα πρέπει να διορθωθούν οι ασάφειες λόγω βάθμωσης και αντιμετάθεσης. Αυτό το δεύτερο στάδιο αναλύεται στην επόμενη ενότητα.

5.5 Ασάφειες λόγω βάθμωσης και αντιμετάθεσης

Ας συμβολίσουμε με $\hat{\mathbf{H}}(f)$ μια εκτίμηση του πίνακα $\mathbf{H}(f)$

$$\hat{\mathbf{H}}(f) = \mathbf{H}(f)\mathbf{D}^{-1}(f)\mathbf{\Pi}^{-1}(f) \quad (5.17)$$

όπου $\mathbf{\Pi}(f)$ είναι ένας άγνωστος αντιμεταθετικός πίνακας και $\mathbf{D}(f)$ ένας άγνωστος διαγώνιος πίνακας. Για να αντιμετωπίσουμε τις ασάφειες λόγω βάθμωσης και αντιμετάθεσης, πρέπει να υπολογίσουμε τα $\mathbf{D}(f)$ και $\mathbf{\Pi}(f)$.

5.5.1 Ασάφεια λόγω βάθμωσης

Μια πιθανή προσέγγιση για να αντιμετωπίσουμε την ασάφεια λόγω βάθμωσης είναι η επονομαζόμενη αρχή της ελάχιστης παραμόρφωσης (*minimal distortion principle* [70], [71]). Επιλέγουμε

$$\mathbf{D}(f) = \text{diag}[\mathbf{Q}\hat{\mathbf{H}}(f)] \quad (5.18)$$

όπου $\mathbf{Q} \in \mathbb{R}^{I \times J}$ είναι ένας πίνακας όλες οι είσοδοι του οποίου είναι $1/J$ και ο $\text{diag}(\cdot)$ διατηρεί μόνο τα διαγώνια στοιχεία και μηδενίζει τα μη-διαγώνια. Αυτή η επιλογή του $\mathbf{D}(f)$ μπορεί να ερμηνευτεί με βάση τα παρακάτω. Αν ο $\hat{\mathbf{H}}(f)$ είναι πλήρης στηλών για κάθε συχνότητα, μπορούμε να ορίσουμε τους πίνακες διαχωρισμού $\hat{\mathbf{W}}(f) \stackrel{\text{def}}{=} \hat{\mathbf{H}}^\dagger(f), f = 1, \dots, F$. Το σύστημα μίξης χαρακτηρίζεται σε συχνότητα f από την ακόλουθη εξίσωση:

$$\mathbf{x}(f, q) = \mathbf{H}(f)\mathbf{s}(f, q) \quad (5.19)$$

Αν πολλαπλασιάσουμε από αριστερά και τις δύο πλευρές της εξίσωσης 5.19 με $\hat{\mathbf{W}}(f)$ προκύπτει:

$$\hat{\mathbf{s}}(f, q) \stackrel{\text{def}}{=} \hat{\mathbf{W}}(f)\mathbf{x}(f, q) = \mathbf{\Pi}(f)\text{diag}[\mathbf{Q}\hat{\mathbf{H}}(f)]\mathbf{s}(f, q) \quad (5.20)$$

Συνεπάγεται ότι

$$\hat{s}_i(f, q) = \frac{1}{J} \sum_{j=1}^J \hat{h}_{j,i}(f) s_{\Pi(i)}(f, q) \quad (5.21)$$

όπου $s_{\Pi(i)}(f, q)$ είναι το i -οστό στοιχείο του $\mathbf{\Pi}(f)\mathbf{s}(f, q)$. Στην περίπτωση τέλει διαχωρισμού, η ερμηνεία της (21) είναι ότι η i -οστή έξοδος αλγορίθμου του τυφλού διαχωρισμού σημάτων είναι ο μέσος όρος όλων των παρατηρήσεων της $\Pi(i)$ ής πηγής μεταξύ των αισθητήρων, όταν όλες οι πηγές είναι κλειστές. Στόχος μας τώρα είναι να υπολογίσουμε τους αντιμεταθετικούς πίνακες $\mathbf{\Pi}(f)$, $f = 1, \dots, F$, ώστε η i -οστή έξοδος $\hat{s}_i(f, q)$ στην (21) να ενώνει μεταξύ τους τα φασματικά στοιχεία που προέρχονται από την ίδια πηγή $s_{\Pi(i)}(f, q)$ κατά μήκος όλων των τμημάτων της συχνότητας.

5.5.2 Ασάφεια λόγω αντιμετάθεσης

Η φασματική ευθυγράμμιση είναι αρκετά δύσκολο πρόβλημα. Αν έχουμε I πηγές, τότε υπάρχουν $I!$ πιθανές αντιμεταθέσεις για κάθε συχνότητα, κάτι που θέτει ένα δύσκολο συνδυαστικό πρόβλημα. Πολλές τεχνικές έχουν προταθεί στη βιβλιογραφία για την επίλυση του προβλήματος της αντιμετάθεσης. Για την έρευνα μας στην παρούσα εργασία παραπέμπουμε στο [51]. Πολλές τεχνικές στηρίζονται στη γεωμετρική πληροφορία, όπως ο υπολογισμός της *Κατεύθυνσης Της Αφίξης (Direction Of Arrival-DOA)* [70]. Άλλες τεχνικές βασίζονται στη συνοχή των συντελεστών φίλτρου. Η τελευταία προσέγγιση εκμεταλλεύεται προηγούμενη γνώση σχετικά με τα φίλτρα μίξης και η λύση μπορεί να επιτευχθεί απαιτώντας η απόκριση συχνότητας $\mathbf{H}(f)$ του φίλτρου μίξης να είναι συνεχής στην f . Είναι επίσης πιθανό να επιβάλλουμε την εξομάλυνση των τιμών των φίλτρων διαχωρισμού στο πεδίο της συχνότητας. Αυτό το επιτυγχάνεται στο [51] περιορίζοντας τις αναβαθμίσεις των φίλτρων διαχωρισμού στο πεδίο της συχνότητας στην εξίσωση 5.2 ώστε να έχουν μια περιορισμένη υποστήριξη στο πεδίο του χρόνου, δηλαδή, $\mathbf{W}(\tau) = 0$ για $\tau > K \ll F$. Ο περιορισμός του μήκους του φίλτρου μπορεί να δημιουργήσει προβλήματα σε περιβάλλοντα με μεγάλη αντήχηση όπου μεγάλα φίλτρα διαχωρισμού είναι απαραίτητα για να λάβουν υπόψη τους όλες τις ανακλάσεις. Στο [51] αναφέρεται ότι αν χρειαζόμαστε ένα μεγάλο φίλτρο διαχωρισμού μήκους K , μπορεί κανείς να διαλέξει ένα μεγάλο πλαίσιο μεγέθους F έτσι ώστε να ισχύει ο περιορισμός $K \ll F$ λόγω της προσέγγισης της κυκλικής συνέλιξης. Ωστόσο, οι μεγάλες τιμές του F αυξάνουν σημαντικά τη συνολική πολυπλοκότητα. Μια άλλη κατηγορία τεχνικών διόρθωσης της αντιμετάθεσης εκμεταλλεύεται τις ιδιότητες των σημάτων ομιλίας. Μια ευρέως χρησιμοποιούμενη ιδιότητα η συσχέτιση των συχνοτήτων των φακέλων των σημάτων ομιλίας [72], [73], που οφείλεται στη φύση της παραγωγής ομιλίας. Για παράδειγμα, όταν ο ομιλητής μιλάει πιο δυνατά, όλες οι φασματικές συνιστώσες του σήματος τείνουν να αυξηθούν σε επίπεδο, και αντιστρόφως. Με βάση αυτή την ιδέα, έχουν προταθεί πολλά κριτήρια και σχετικές στρατηγικές για να επιβάλλουν τη σύζευξη της συχνότητας μεταξύ των γειτονικών τμημάτων συχνότητας [50],[74]. Το μεγαλύτερο μειονέκτημα των διαδοχικών στρατηγικών προσαρμογής είναι η *αναπαραγωγή σφάλματος*, δηλαδή, ένα σφάλμα που έγινε στη διόρθωση αντιμετάθεσης στη συχνότητα f μπορεί να επηρεάσει σημαντικά τη διόρθωση στις ακόλουθες συχνότητες. Για να αποφύγουμε αυτό το πρόβλημα, μια πιθανή προσέγγιση είναι να χρησιμοποιήσουμε μια μέθοδο ομαδοποίησης για να υπολογίσουμε ένα προφίλ

αναφοράς (ή κεντροειδές) για κάθε διαχωρισμένη πηγή το οποίο θα είναι ανεξάρτητο από τη συχνότητα, και στη συνέχεια να αντιμετωπίσουμε, για κάθε συχνότητα, τα I προφίλ τα οποία εξαρτώνται από τη συχνότητα με τέτοιο τρόπο ώστε όλα να συνδέονται σε ένα διαφορετικό προφίλ αναφοράς. Τα τρία στοιχεία κλειδιά της τεχνικής της ομαδοποίησης είναι τα ακόλουθα:

1. Ο ορισμός των ποσοτήτων που ομαδοποιούνται, δηλαδή, το προφίλ των πηγών.
2. Το μέτρο που χρησιμοποιείται για να ποσοτικοποιήσει το αντίστοιχο επίπεδο μεταξύ των κεντροειδών και των προφίλ (π.χ. συσχέτιση, απόσταση, κτλ.).
3. Η στρατηγική ομαδοποίησης.

Στο [54], το προφίλ $\hat{y}_i(f, q)$ ενός διαχωρισμένου σήματος $\hat{s}_i$ θεωρείται ότι είναι ο φάκελός του, $\hat{y}_i(f, q) = |\hat{s}_i(f, q)|$. Στο [55], το προφίλ $\hat{y}_i(f, q)$ είναι ένα συγκεκριμένο μέτρο κυριαρχίας. Στο [52], το προφίλ της i -ιστής διαχωρισμένης πηγής ορίζεται από την κεντροθετημένη λογαριθμική φασματική πυκνότητα ισχύος $\hat{y}_i(f, q) = \log[\hat{W}_{i,:}(f) \mathbf{R}_x(f, q) \hat{W}_{i,:}^H(f)]$. Το μήκος N_f των προφίλ είναι μια εξίσου σημαντική παράμετρος για την ακρίβεια της προσέγγισης που βασίζεται στην ομαδοποίηση, ειδικά για μικρά σήματα. Στην πράξη, τα $\hat{y}_i(f, q)$ προφίλ υπολογίζονται για επικαλυπτόμενα πλαίσια σε ολόκληρο το σήμα. Μόλις υπολογιστούν τα προφίλ, επόμενο βήμα είναι να υπολογιστούν τα κεντροειδή και να εκτελεστεί η ομαδοποίηση. Η βασική υπόθεση των προσεγγίσεων που στηρίζονται στην ομαδοποίηση είναι ότι τα προφίλ που προέρχονται από την ίδια πηγή, αλλά από διαφορετικές συχνότητες, είναι πολύ πιο όμοια από αυτά που προέρχονται από διαφορετικές πηγές. Για να συσχετίσουμε κάθε προφίλ πηγής με ένα κεντροειδές για κάθε συχνότητα, μπορεί κανείς να μεγιστοποιήσει τα μέτρα συσχέτισης [54], [55] ή να ελαχιστοποιήσουμε τα μέτρα της απόστασης [52] σε όλες τις I πιθανές αντιμετωπίσεις για κάθε συχνότητα. Σε αυτό το σημείο, η στρατηγική της ομαδοποίησης είναι σημαντική. Στα [52], [54] και [55], τα κεντροειδή και οι πίνακες αντιμετάθεσης αναβαθμίζονται με έναν επαναληπτικό τρόπο. Για κάθε επανάληψη, τα κεντροειδή αρχικά αναβαθμίζονται ως ο μέσος όρος όλων των συχνοτήτων των τρεχόντων προφίλ των πηγών. Στη συνέχεια, τα προφίλ των πηγών αντιμετωπίζονται έτσι ώστε να ταιριάζουν με τα τρέχοντα κεντροειδή, σύμφωνα με το επιλεγμένο μέτρο (απόσταση στο [52] ή συσχέτιση στο [54] και στο [55]). Ωστόσο, ο υπολογισμός αυτού του μέτρου για I αντιμετωπίσεις και F συχνότητες σε κάθε επανάληψη συμπεριλαμβάνει ένα σημαντικό υπολογιστικό κόστος.

Σε αυτή την ενότητα, προτείνουμε μια πιο επαρκή στρατηγική ομαδοποίησης για να αποφύγουμε αυτό το πρόβλημα. Αντίθετα με τις προαναφερθείσες πλήρως επαναληπτικές στρατηγικές, δεν παρεμβάλλονται οι αναβαθμίσεις των κεντροειδών και των πινάκων αντιμετάθεσης, κάτι που μειώνει σημαντικά την πολυπλοκότητα. Το σύστημά μας μπορεί να συνοψιστεί ως εξής:

1ο Βήμα. Υπολογισμός των κεντροειδών: Θεωρούμε τον $I \times N_f$ πίνακα $\hat{\mathbf{T}}(f)$ που συλλέγει τα I $\hat{y}_i(f)$, $i = 1, \dots, I$ προφίλ. Ο $FI \times N_f$ πίνακας $\hat{\mathbf{T}}$ απορρέει από την αλληλουχία των πινάκων $\hat{\mathbf{T}}(f)$, $f = 1, \dots, F$. Αφού τα προφίλ έχουν υπολογιστεί για επικαλυπτόμενα πλαίσια, ο $\hat{\mathbf{T}}$ διατηρεί ένα σύνολο FI σημείων ομαλά μεταβαλλόμενων με τον χρόνο. Στόχος μας τώρα είναι να διαμοιράσουμε

αυτά τα σημεία σε I ομάδες. Αυτό μπορεί να γίνει με τη χρήση του k -αλγορίθμου στον $\hat{\mathbf{I}}$, γεγονός που παράγει έναν $I \times N_f$ κεντροειδή πίνακα $\mathbf{M} = [\mathbf{m}_1^T, \dots, \mathbf{m}_I^T]^T$. –το άθροισμα της ομάδας των αποστάσεων σημείου-κέντρου βάρους ελαχιστοποιείται. Αυτός ο κεντροειδής πίνακας είναι τέτοιος ώστε το άθροισμα σε όλες τις ομάδες, των εντός των ομάδων αθροισμάτων των αποστάσεων σημείου-κεντροειδούς να ελαχιστοποιείται.

2^ο Βήμα. Βρίσκοντας τους αντιμεταθετικούς πίνακες: Για κάθε τμήμα της συχνότητας, ψάχνουμε τον $I \times I$ ανιμεταθετικό πίνακα $\mathbf{\Pi}(f)$ τέτοιο ώστε ο $\hat{\mathbf{I}}(f) \mathbf{\Pi}(f)$ να ταιριάζει με τον $\mathbf{M}$, σύμφωνα με το επιλεγμένο μέτρο. Μια πιθανή επιλογή [52] είναι να λύουμε την

$$\min_{\mathbf{\Pi}(f)} \varphi(f), f = 1, \dots, F \quad (5.22)$$

Πίνακας 5.1

	Κριτήριο K1	Κριτήριο K2 και K3
Μέθοδος ομαδοποίησης	Προφίλ λογαριθμικής ενέργειας με ένα μέτρο απόστασης [8]	K2: Κυρίαρχα προφίλ με ένα μέτρο συσχέτισης [22]
		K3: Προφίλ φάκελοι με ένα μέτρο συσχέτισης [21]
Επανάληψη	$O(FN_f I^2 (I-1)! n)$	$O(FN_f i(I+1)n + FI^2(I-1)! n)$
k-μέσο	$O(FN_f I^2 n + FN_f I^2 (I-1)! n)$	$O(FN_f I^2 n + FN_f I^2 + FI^2(I-1)!)$

όπου $\varphi(f) \stackrel{\text{def}}{=} \|\mathbf{M} - \hat{\mathbf{I}}(f) \mathbf{\Pi}(f)\|_F^2$. Μια άλλη εναλλακτική [21], [22] είναι να λύσουμε την

$$\max_{\mathbf{\Pi}(f)} \sum_{i=1}^I \rho(m_i, [\hat{\mathbf{I}}(f) \mathbf{\Pi}(f)]_{:,i}) \quad (5.23)$$

όπου το ρ είναι ο συντελεστής συσχέτισης. Για να λύσουμε την (5.22) ή την (5.23), υπολογίζουμε την πλήρη σειρά των μέτρων $I!$ για κάθε συχνότητα και διατηρούμε τον αντιμεταθετικό πίνακα που οδηγεί στην καλύτερη λύση.

Το κύριο χαρακτηριστικό στο σύστημα μας είναι ότι μόνο το Βήμα 1 είναι επαναληπτικό και η (5.22) ή η (5.23) έχει λυθεί μόνο μία φορά. Πρόκειται για ένα σημαντικό πλεονέκτημα έναντι των εξ ολοκλήρου επαναληπτικών στρατηγικών που χρησιμοποιούνται στο [52], [54], [55], όπου η (5.22) ή η (5.23) επιλύονται σε κάθε επανάληψη.

5.5.3 Σύγκριση μεταξύ των λυτών αντιμετάθεσης

Στην ενότητα αυτή, θα συγκρίνουμε την πολυπλοκότητα και την απόδοση των ακόλουθων κριτηρίων για να λύσουμε το πρόβλημα της αντιμετάθεσης: (K1) ομαδοποίηση των προφίλ λογαριθμικής ενέργειας με μέτρο απόστασης (5.22), όπως προτείνεται στο [52], (K2) ομαδοποίηση των

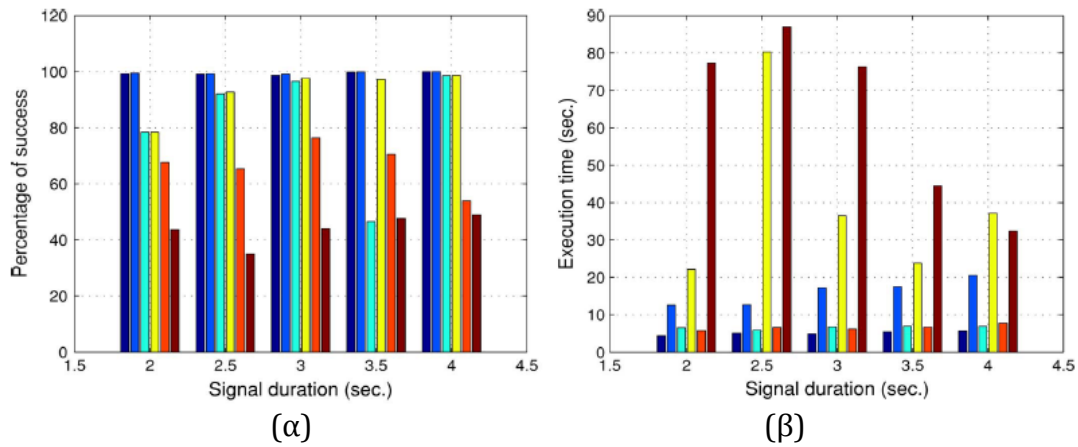
κυρίαρχων προφίλ με ένα μέτρο συσχέτισης (5.23), όπως προτείνεται στην [55], (K3) ομαδοποίηση των φακέλων-προφίλ με μέτρο συσχέτισης (5.23), όπως προτείνεται στο [54]. Τα κριτήρια αυτά συνδυάζονται είτε με μια εξ ολοκλήρου επαναληπτική στρατηγική ομαδοποίησης, όπως στην αρχική τους εκδοχή, ή με την προσέγγιση με k -μέσα που προτείναμε. Στον Πίνακα 5.1 αναφέρονται οι εντολές πολυπλοκότητας των διαφόρων συνδυασμών. Είναι σαφές ότι η στρατηγική ομαδοποίησης που προτείναμε έχει χαμηλότερη πολυπλοκότητα από τον πλήρως επαναληπτικό ομόλογό του. Αυτό προκύπτει από το πλεονέκτημα του να υπολογίσουμε μόνο τα κεντροειδή με έναν επαναληπτικό τρόπο, αντί να παρεμβάλλουμε ενημερώσεις των κεντροειδών και των αντιμεταθετικών πινάκων.

Στο Σχήμα 5.1, συγκρίνουμε την απόδοση των διαφόρων αντιμεταθετικών λυτών που εφαρμόζονται σε αυθαίρετα αντιμεταθετημένες εκδοχές των προφίλ πραγματικών πηγών $\mathbf{I}(f)$, δηλαδή, προσομοιώνουμε την έξοδο ενός τέλει σταδίου διαχωρισμού. Η ανεξάρτητη από τη συχνότητα αντιμετάθεση που απομένει επιλύεται με μια διαδικασία αντιστοίχισης στήλης, μετά την οποία υπολογίζουμε τον αριθμό των συχνοτήτων για τις οποίες οι $\hat{\mathbf{I}}(f)$ είναι απόλυτα ευθυγραμμισμένοι και υπολογίζουμε το ποσοστό της επιτυχίας. Το τελευταίο αναπαρίσταται στο Σχήμα 5.1 για $I = 5$ πηγές. Ο συνολικός χρόνος εκτέλεσης φαίνεται επίσης. Από αυτό το σχήμα, είναι σαφές ότι η ομαδοποίηση των προφίλ λογαριθμικής ενέργειας φαίνεται να είναι μια πολύ αποδοτική λύση για να λυθεί το πρόβλημα της αντιμετάθεσης, αφού η απόδοση είναι κοντά στο 100%, ακόμη και με πέντε πηγές των 2s μόνο. Συγκριτικά, τα δύο άλλα κριτήρια (κυρίαρχα προφίλ και προφίλ-φακέλου) είναι πιο ευαίσθητα ως προς το μήκος του σήματος. Όπως αναμένεται, ο συνδυασμός των k -μέσων που βασίζεται στη στρατηγική ομαδοποίησης με τα τρία κριτήρια επιτρέπει μια πολύ σημαντική μείωση της πολυπλοκότητας, σχετική με την απόλυτα επαναληπτική προσέγγιση. Με βάση αυτές τις παρατηρήσεις, καθώς η ομαδοποίηση των προφίλ λογαριθμικής ενέργειας με μια στρατηγική βασισμένη στο k -μέσο προσφέρει την καλύτερη συναλλαγή μεταξύ πολυπλοκότητας και απόδοσης, θα χρησιμοποιήσουμε αυτό το κριτήριο μετά το PARAFAC στάδιο διαχωρισμού σε πραγματικές καταστάσεις τυφλού διαχωρισμού.

5.6 Υπο-ορισμένο πρόβλημα

Αν $\hat{\mathbf{H}}(f)$ είναι πλήρως-στήλης βαθμού για κάθε συχνότητα, ο διαχωρισμός μπορεί να επιτευχθεί στο πεδίο της συχνότητας από το $\hat{\mathbf{s}}(f, q) = \hat{\mathbf{W}}(f)\mathbf{x}(f, q)$, όπου ο $\hat{\mathbf{W}}(f) = \hat{\mathbf{H}}^\dagger(f)$ λαμβάνεται μετά τη διόρθωση των ασαφειών της βάρθρωσης και της αντιμετάθεσης. Οι διαχωρισμένες πηγές στη συνέχεια υπολογίζονται εφαρμόζοντας τον αντίστροφο διακριτό μετασχηματισμό Fourier (Inverse DFT) στο $\{\hat{\mathbf{s}}(f, q), f = 1, \dots, F\}$. Εναλλακτικά, κάποιος μπορεί πρώτα να υπολογίσει τον $\hat{\mathbf{W}}$ πίνακα-φίλτρο διαχωρισμού στο πεδίο του χρόνου, λαμβάνοντας τον αντίστροφο DFT του $\{\hat{\mathbf{W}}(f), f = 1, \dots, F\}$, μετά από το οποίο η διαδικασία της συνέλιξης (5.2) μπορεί να υπολογιστεί αποτελεσματικά μέσω μιας διαδικασίας προσθετικής-επικάλυψης. Η τελευταία αυτή προσέγγιση θα χρησιμοποιηθεί στην πράξη.

Στην υπο-ορισμένη περίπτωση, το πρόβλημα είναι πιο δύσκολο. Σύμφωνα με τους όρους της μοναδικότητας που αναφέρονται στην ενότητα 5.3.2, ο PARAFAC επιτρέπει τον προσδιορισμό του $\mathbf{H}(f)$ με ένα μοναδικό τρόπο, μέχρι να προκύψουν οι ασάφειες βάθμωσης και αντιμετάθεσης. Οι τελευταίες είναι διορθωμένες όπως εξηγείται στην ενότητα 5.5. Ωστόσο, ο προκύπτον πίνακας $\hat{\mathbf{H}}(f)$ δεν είναι πλέον ψευδοαντίστροφος και ως εκ τούτου ο τέλειος διαχωρισμός δεν είναι δυνατός. Σε αυτή την ενότητα, θα δείξουμε ότι η σημαντική μείωση του crosstalk εξακολουθεί να είναι δυνατή με τη χρήση μεθόδων επεξεργασίας πινάκων, ιδίως σε μια χρονικά μεταβαλλόμενη εκδοχή του σχεδιασμού Capon. Πρώτον, παρατηρούμε ότι για ένα αρκετά μικρό υποσύνολο p , η πιθανότητα όλες οι πηγές να έχουν ταυτόχρονα υψηλή φασματική πυκνότητα ισχύος είναι χαμηλή.



Σχήμα 5.1. Απόδοση των τριών κριτηρίων $K1$, $K2$ και $K3$ για την επίλυση του προβλήματος αντιμετάθεσης, συνδυασμένα είτε με τη στρατηγική ομαδοποίησης με ενός βήματος k -μέσου είτε με την πλήρως επαναληπτική στρατηγική. Σε κάθε σχήμα, υπάρχουν πέντε ομάδες, που καθεμία περιλαμβάνει έξι ράβδους. Κάθε ομάδα αντιστοιχεί σε μια συγκεκριμένη διάρκεια σήματος (2, 2.5, 3, 3.5, ή 4s). Μέσα σε κάθε ομάδα, οι ετικέτες από αριστερά προς τα δεξιά είναι οι εξής. (1) $K1$ με k -μέσο. (2) $K1$ επαναληπτική. (3) $K2$ με k -μέσο. (4) $K2$ επαναληπτική. (5) $K3$ με k -μέσο. (6) $K3$ επαναληπτική. (α) Ποσοστό επιτυχίας, $I = 5$ πηγές, $F = 2048$. (β) χρόνος CPU, $I = 5$ πηγές, $F = 2048$.

Για παράδειγμα, αν οι $I - J$ πηγές μεταξύ του I έχουν μια μακρά περίοδο παύσης μέσα σε ένα υποσύνολο p , το υπο-ορισμένο πρόβλημα σχεδόν συνεχίζει για αυτό το υποσύνολο σε ένα $J \times J$ καθορισμένο πρόβλημα. Αυτό υποδηλώνει ότι η μείωση του crosstalk θα πρέπει να πραγματοποιείται σε μια ανά υποσύνολο βάση, για να λογοδοτεί για τις μεταβολές στις ισχύς του crosstalk (σημειώνεται ότι η μέθοδος μας αυτόματα προσαρμόζεται σε αυτές τις μεταβολές· δεν απαιτεί ανίχνευση λειτουργίας / παύσης). Στόχος μας είναι στη συνέχεια να βρούμε ένα σύνολο από πίνακες διαχωρισμού $\{\hat{\mathbf{W}}(f, p), f = 1, \dots, F, p = 1, \dots, P\}$, τέτοιοι ώστε να μειώνεται το crosstalk για κάθε συχνότητα και για κάθε υποσύνολο. Αυτό μπορεί να επιτευχθεί με σχεδιασμό Capon. Για μία δεδομένη πηγή i , ένα δεδομένο σύνολο p και μια δεδομένη συχνότητα f , ψάχνουμε ένα $J \times 1$ σχεδιασμό διάνυσμα $\hat{\mathbf{w}}_i(f, p)$ έτσι ώστε το

$$\begin{aligned}\hat{s}_i(f, p) &= \hat{\mathbf{w}}_i^H(f, p)x(f, p) \\ &= \hat{\mathbf{w}}_i^H(f, p)\hat{\mathbf{h}}_i(f)s_i(f, p) + \sum_{k \neq i} \hat{\mathbf{w}}_i^H(f, p)\hat{\mathbf{h}}_k(f)s_k(f, p)\end{aligned}\quad (5.24)$$

να διατηρεί τον πρώτο όρο και να καταστέλλει το δεύτερο. Εδώ, το $\hat{\mathbf{h}}_i(f)$ συμβολίζει την i -στή στήλη του $\hat{\mathbf{H}}(f)$ μετά τη διόρθωση των ασαφειών βάθμωσης και αντιμετάθεσης. Στην (5.24), το $\hat{s}_i(f, p)$ προέρχεται από το άθροισμα ενός σήματος ενδιαφέροντος και σημάτων crosstalk. Το διάνυσμα $\hat{\mathbf{w}}_i(f, p)$, που ελαχιστοποιεί το SIR του σχεδιασμού Capon που λύνει την

$$\begin{aligned}\min_{\hat{\mathbf{w}}_i(f, p)} & \hat{\mathbf{w}}_i^H(f, p)\mathbf{R}_x(f, p)\hat{\mathbf{w}}_i(f, p) \\ \text{s. t. } & \hat{\mathbf{w}}_i^H(f, p)\hat{\mathbf{h}}_i(f) = 1\end{aligned}\quad (5.25)$$

Η λύση αυτού του προβλήματος είναι

$$\hat{\mathbf{w}}_i(f, p) = \frac{\mathbf{R}_x^{-1}(f, p)\hat{\mathbf{h}}_i(f)}{\hat{\mathbf{h}}_i^H(f)\mathbf{R}_x^{-1}(f, p)\hat{\mathbf{h}}_i(f)}\quad (5.26)$$

Στη συνέχεια εφαρμόζεται ο σχεδιασμός Capon σε κάθε συχνότητα για κάθε πηγή και για κάθε υποσύνολο.

5.7 Υλοποίηση με συνεχή επαναπροσδιορισμό

Στις προηγούμενες ενότητες, θεωρήσαμε ένα συνεχές περιβάλλον μίξης και προτείναμε μια συγκεκριμένη υλοποίηση του PARAFAC (Batch PARAFAC) στο πεδίο της συχνότητας. Ωστόσο, σε πραγματικές καταστάσεις, το σύστημα μίξης μπορεί να θεωρηθεί ως συνεχές μόνο για σύντομα χρονικά διαστήματα, λόγω της κινητικότητας του ομιλητή, στις διακυμάνσεις τους περιβάλλοντος, κλπ. Οι προσαρμοστικοί αλγόριθμοι του τυφλού διαχωρισμού σημάτων με συνεχή επαναπροσδιορισμό έχουν ως εκ τούτου μεγάλο ενδιαφέρον [75], [76]. Σε αυτή την ενότητα, θα δείξουμε ότι η προσαρμογή του τυφλού διαχωρισμού σημάτων βασισμένη στον batch PARAFAC στην περίπτωση του συνεχούς επαναπροσδιορισμού μπορεί να μετατραπεί σε πρόβλημα παρακολούθησης ενός μόνο διαχωρισμού PARAFAC για κάθε συχνότητα, για την οποία έχουμε πρόσφατα προτείνει αποτελεσματικούς προσαρμοστικούς αλγόριθμους [56].

Ας ξεκινήσουμε με την εξίσωση 5.14, η οποία αντιπροσωπεύει το μοντέλο PARAFAC του διανύσματος αυτοσυσχέτισης εξόδου $\mathbf{R}_x(f) \in \mathbb{C}^{J \times J \times P}$, με αναπαράσταση πίνακα $\mathbf{R}_x(f) \in \mathbb{C}^{J^2 \times P}$. Εάν ο πίνακας μίξης $\mathbf{H}(f)$ κυμαίνεται μεταξύ δύο διαδοχικών διαστημάτων, πρέπει να αναπροσαρμοστεί από το χρόνο και ο παρατηρούμενος πίνακας αυτοσυσχέτισης είναι τώρα

$$\mathbf{R}_x(f) = [(\mathbf{H}(f, 1) \odot \mathbf{H}^*(f, 1))c^T(f, 1), \dots, (\mathbf{H}(f, P) \odot \mathbf{H}^*(f, P))c^T(f, P)]\quad (5.27)$$

όπου $c^T(f, P)$ είναι η p -οστή στήλη του $\mathbf{C}^T(f)$. Κατά συνέπεια, το μοντέλο PARAFAC, και αντίστοιχα η JAD διατύπωση, παραμένουν περίπου έγκυρα μόνο

αν η ο πίνακας μίξης $\mathbf{H}(f, P)$ είναι σχεδόν σταθερός κατά τη διάρκεια των P διαδοχικών χρονικών καθυστερήσεων. Για ένα επαρκώς σύντομο χρονικό διάστημα $L_k = [t_k: t_{P+k-1}]$, που αποτελείται από P διαδοχικά χρονικά σύνολα, μπορούμε να γράψουμε

$$\mathbf{R}_x(f, L_k) \simeq [\mathbf{H}(f, L_k) \odot \mathbf{H}^*(f, L_k) \mathbf{C}^T(f, L_k)] \quad (5.28)$$

όπου $\mathbf{H}(f, L_k) \simeq \mathbf{H}(f, k) \simeq \dots \simeq \mathbf{H}(f, P + k - 1)$ και $\mathbf{C}^T(f, L_k) = [c^T(f, k), \dots, c^T(f, P + k - 1)]$.

Το πρόβλημα μπορεί να συνοψιστεί ως εξής:

«Οι δοσμένες εκτιμήσεις των $\mathbf{H}(f, L_k)$ και $\mathbf{C}(f, L_k)$, υπολογίζουν τα $\mathbf{H}(f, L_{k+1})$ και $\mathbf{C}(f, L_{k+1})$ από τους παρατηρούμενους πίνακες $\mathbf{R}_x(f, L_k)$ και $\mathbf{R}_x(f, L_{k+1})$ »

Μία πιθανή λύση στο πρόβλημα αυτό είναι να εφαρμόσουμε επανειλημμένα έναν batch PARAFAC αλγόριθμο στα διαδοχικά σύντομα χρονικά διαστήματα L_k . Αν και ο αλγόριθμος PARAFAC-SD αποδείχθηκε πολύ γρήγορος συγκρινόμενος με τις υπάρχουσες τεχνικές διαγωνιοποίησης, η προσαρμοσμένη του εκδοχή θα ήταν πολύ επιθυμητή. Αυτή ακριβώς είναι η ουσία του PARAFAC-SDT ("PARAFAC μέσω Παρακολούθησης της Ταυτόχρονης Διαγωνιοποίησης") αλγόριθμος που προτείνεται στο [56]. Ο PARAFAC-SDT λύνει (5.16) προσαρμοστικά, παρακολουθώντας αρχικά το SVD του $\mathbf{R}_x(f)$ πριν ενημερώσει αναδρομικά τα $\mathbf{Z}(f)$ και $\mathbf{H}(f)$. Για περισσότερες λεπτομέρειες σχετικά με αυτόν τον αλγόριθμο, παραπέμπουμε στο [56].

Κατ' αρχήν, ένας λύτης προσαρμοστικής αντιμετάθεσης είναι επίσης απαραίτητος για να καταλήξουμε σε μια ολοκληρωμένη λύση προσαρμοστικού τυφλού διαχωρισμού σημάτων. Ευτυχώς, μια ωφέλιμη πλευρά της παρακολούθησης με χρήση του PARAFAC-SDT είναι ότι οι ενημερώσεις αυξάνονται εγγενώς- διατηρώντας έτσι φυσικά τη σωστή αντιμετάθεση, με την προϋπόθεση ότι ο προσαρμοστικός αλγόριθμος έχει προετοιμαστεί σωστά. Τέλος, υπάρχουν προσαρμοστικές εφαρμογές του σχεδιασμού Caron, και αυτές μπορούν εύκολα να τροποποιηθούν για να δημιουργήσουν μια πλήρη λύση με συνεχή επαναπροσδιορισμό που μπορεί να εφαρμοστεί στις υπο-ορισμένες περιπτώσεις.

Κεφάλαιο 6

Εφαρμογή του μοντέλου PARAFAC-Μετρήσεις

6.1 Εισαγωγή

Στο κεφάλαιο αυτό πραγματοποιούμε μετρήσεις χρησιμοποιώντας το μοντέλο PARAFAC. Αρχικά, μελετάμε το πρόβλημα του υπερ-ορισμένου διαχωρισμού σημάτων για την περίπτωση της συνελκτικής μείξης των σημάτων ταυτόχρονων ομιλητών σε περιβάλλον πραγματικού δωματίου. Αρχικά, μελετάμε αν η χρήση περισσότερων μικροφώνων αυξάνει την πληροφορία που λαμβάνεται από τα μικρόφωνα. Το γεγονός αυτό θα μπορούσε να βελτιώσει την απόδοση των αλγορίθμων διαχωρισμού και κατά συνέπεια να συμβάλει και στην αύξηση της αξιοπιστίας των χρησιμοποιούμενων συστημάτων αναγνώρισης ομιλίας. Επιπλέον, μελετάμε πώς ο αριθμός των εικονικών πηγών, η μεταβολή του συντελεστή ανακλάσεως και η μείωση της απόστασης πηγών-μικροφώνων επηρεάζει την απόδοση του αλγορίθμου PARAFAC. Στη συνέχεια, θα ασχοληθούμε με το υπο-ορισμένο πρόβλημα διαχωρισμού και θα δείξουμε πώς η μεταβολή των αποστάσεων των μικροφώνων από τις πηγές καθώς και η μεταβολή των διαστάσεων του δωματίου επηρεάζουν την ποιότητα του διαχωρισμού.

Οφείλουμε να σημειώσουμε στο σημείο αυτό ότι οι κώδικες που χρησιμοποιήθηκαν για τη διεκπεραίωση των μετρήσεων υπάρχουν στο paper: *D. Nion, K. N. Mokios, N. D. Sidiropoulos and A. Potamianos, "Batch and Adaptive PARAFAC-based Blind Separation of Convolutional Speech Mixtures", IEEE Trans. on Audio, Speech and Language Processing, Vol. 18, No. 6, pp. 1193-1207, August 2010. Ο αλγόριθμος που χρησιμοποιήθηκε για την διεκπεραίωση των μετρήσεων παρατίθεται στο Παράρτημα.*

6.2 Διαδικασία υλοποίησης του αλγορίθμου της συνελκτικής μίξης

Έστω ότι έχουμε I πηγές και J μικρόφωνα, τα βήματα που πραγματοποιούνται για τη συνέλιξη είναι τα ακόλουθα:

1. Ο μετασχηματισμός Fourier εφαρμόζεται σε διαδοχικά επικαλυπτόμενα πλαίσια έτσι ώστε να πάρουμε ένα στιγμιαίο πρόβλημα τυφλού διαχωρισμού σημάτων για κάθε συχνότητα. Έπειτα, υπολογίζεται ένα σύνολο πινάκων συνδιασποράς για κάθε συχνότητα, για να πάρουμε ένα πρόβλημα από κοινού διαγωνιοποίησης, για κάθε συχνότητα.

2. Για κάθε συχνότητα, το κλασικό πρόβλημα από κοινού διαγωνιοποίησης λύνεται με τον αλγόριθμο διαχωρισμού candecomp/parafac (CP) του τανυστή που έχει κατασκευαστεί από τη συσσώρευση των πινάκων συνδιασποράς σχετικών με αυτή τη συχνότητα στον τρίτο τρόπο λειτουργίας.
3. Για κάθε συχνότητα, οι εκτιμώμενες πηγές φασματικών συνιστωσών θεωρούνται αυθαίρετα αντιμετωπιζόμενες. Αντιμετωπίζουμε αυτό το εξαρτημένο από τη συχνότητα πρόβλημα αντιμετάθεσης είτε ομαδοποιώντας τα εκτιμώμενα προφίλ λογαριθμικής ενέργειας των πηγών είτε ομαδοποιώντας τους λόγους ενέργειας (μέτρο κυριαρχίας).
4. Μόλις λυθεί το εξαρτημένο από τη συχνότητα πρόβλημα αντιμετάθεσης, ακολουθούμε την εξής διαδικασία:

(α) στην περίπτωση του υπερ-ορισμένου προβλήματος, εφαρμόζεται ο αντίστροφος μετασχηματισμός Fourier για να πάρουμε τα διαχωρισμένα φίλτρα στο πεδίο του χρόνου και υπολογίζουμε το διαχωρισμό.

(β) στην περίπτωση του υπο-ορισμένου προβλήματος, χρησιμοποιούμε τον σχεδιασμό Capon για τη μείωση του cross-talk για κάθε πλαίσιο χρόνου-συχνότητας και εφαρμόζουμε τον ifft για κάθε πλαίσιο.

6.3 Αξιολόγηση της απόδοσης

Η μέτρηση της ποιότητας του διαχωρισμού μπορεί να γίνει μέσω του υπολογισμού του λόγου του σήματος προς παρεμβολή (SIR) τόσο των σημάτων που λαμβάνονται στους δέκτες όπου ο λόγος του σήματος προς παρεμβολή όπως θα περίμενε κάποιος είναι πολύ χαμηλός, όσο και των διαχωρισμένων σημάτων. Επιτυχημένος διαχωρισμός των σημάτων εξασφαλίζει υψηλό λόγο σήματος προς παρεμβολή.

Από την εξίσωση 5.2, οι διαχωρισμένες πηγές δίνονται από την

$$\hat{s}_i(t) = \sum_{j=1}^J W_{ij} * x_j(t) \quad (6.1)$$

Το SIR εξόδου για το $\hat{s}_i(t)$ ορίζεται ως ο λόγος της δύναμης του μέρους του $\hat{s}_i(t)$ που έρχεται από την πηγή i , $\hat{s}_{ii}(t)$, προς τη δύναμη των σημάτων του cross-talk $\hat{s}_{ik}(t)$ [7]:

$$SIR_i = 10 \log \frac{\sum_t \hat{s}_{ii}^2(t)}{\sum_t \sum_{k \neq i} \hat{s}_{ik}^2(t)} \quad (6.2)$$

Στα πειράματα αυτού του κεφαλαίου, θα πραγματοποιήσουμε τη συνέλιξη σημάτων ομιλίας με RIRs ήδη μετρημένα σε πραγματικό περιβάλλον ή τεχνητά κατασκευασμένα, έχοντας έτσι πρόσβαση στα σήματα μικροφώνου $x_{ji}(t), j = 1, \dots, J$, τα οποία καταγράφονται όταν η I -οστή πηγή είναι παρούσα. Έτσι, υπολογίζουμε το SIR της I -οστής πηγής ως εξής:

$$SIR_i = 10 \log \frac{\sum_t (\sum_{j=1}^J W_{ij} * x_{ji}(t))^2}{\sum_t \sum_{k \neq i} (\sum_{j=1}^J W_{ik} * x_{jk}(t))^2} \quad (6.3)$$

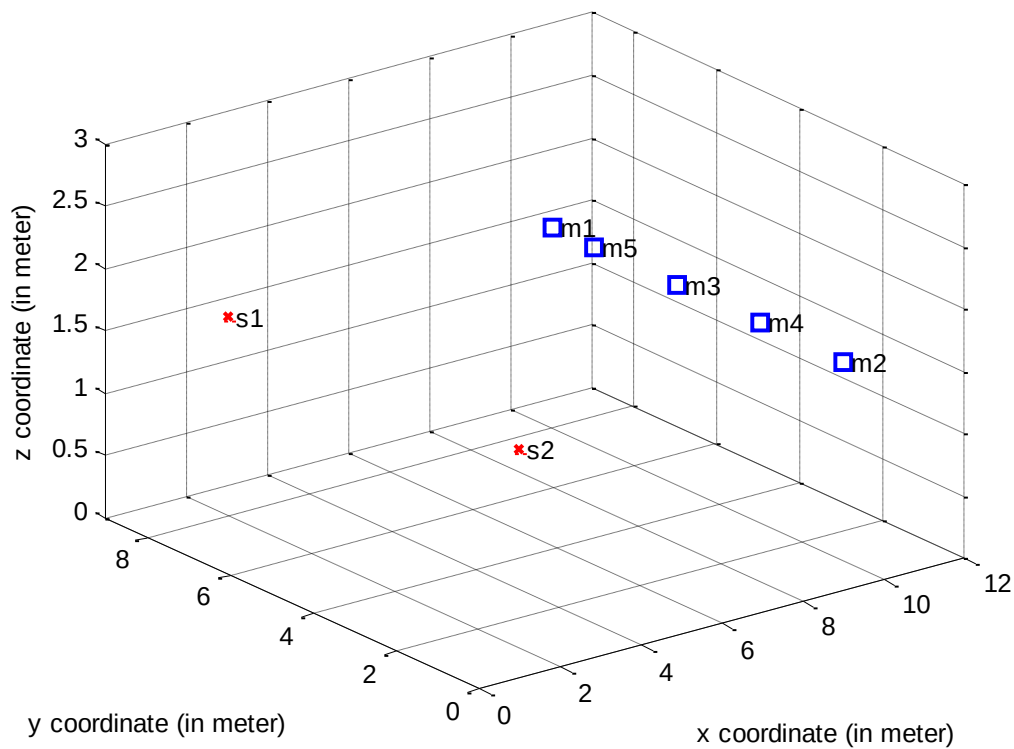
Στην περίπτωση του υπο-ορισμένου διαχωρισμού σημάτων όπου χρησιμοποιείται ο σχεδιασμός Capon σε μια ανά υποσύνολο βάση, το ανάστροφο φίλτρο ποικίλει μεταξύ των υποσυνόλων. Σε αυτή την περίπτωση, το SIR_i υπολογίζεται με έναν παρόμοιο τρόπο, με τη διαφορά ότι τα $\hat{s}_{ii}(t)$ και $\hat{s}_{ik}^2(t)$ στην (6.3) κατασκευάζονται από τη συνένωση των διαδοχικά υπολογισμένων υποσυνόλων.

Ο αλγόριθμος υπολογισμού του SIR στο περιβάλλον του Matlab δίνεται στο Παράρτημα.

6.4 Υπερ-ορισμένος διαχωρισμός σημάτων σε συνελικτική μίξη

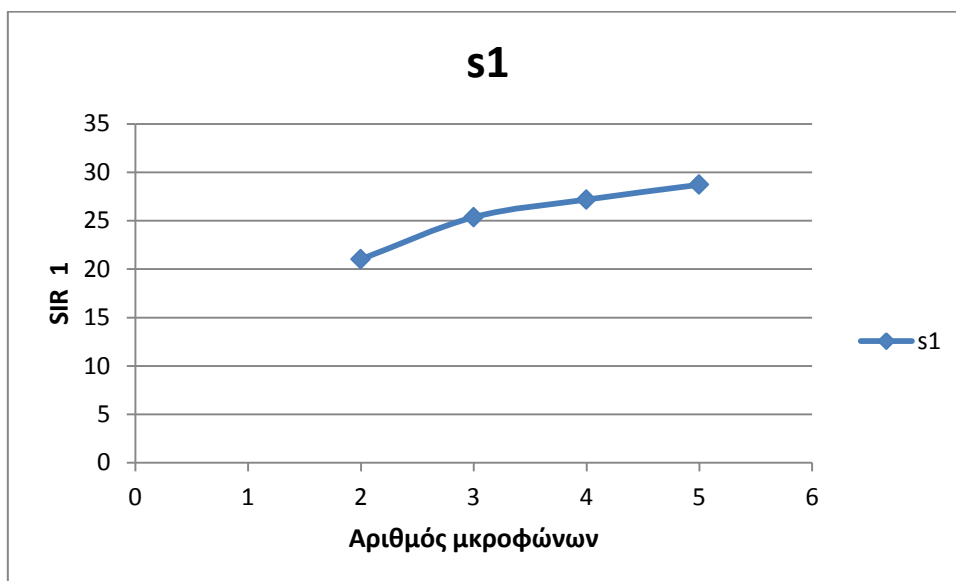
Στην ενότητα αυτή θα μελετήσουμε το πρόβλημα του υπερ-ορισμένου διαχωρισμού σημάτων σε συνελικτική μίξη. Αρχικά, θα μελετήσουμε αν η αύξηση του αριθμού των μικροφώνων στη συνελικτική μίξη βελτιώνει την απόδοση. Οι μετρήσεις πραγματοποιούνται σε ένα δωμάτιο με ανακλάσεις. Ο αριθμός των πηγών είναι $I = 2$ και παίρνουμε μετρήσεις για αριθμό μικροφώνων $J = 2, 3, 4, 5$ αντίστοιχα. Η απεικόνιση του δωματίου φαίνεται στο Σχήμα 6.1.

Spatial configuration : localization of the sources and sensors in the room

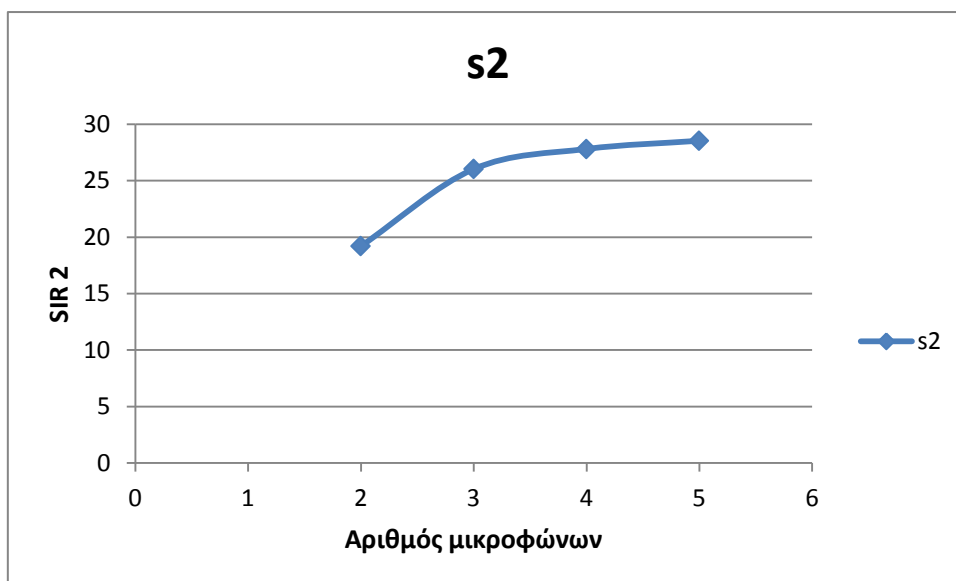


Σχήμα 6.1 Απεικόνιση των δύο πηγών και των μικροφώνων μέσα σε ένα δωμάτιο με ανακλάσεις.

Η βελτίωση της ποιότητας του διαχωρισμού φαίνεται στα παρακάτω σχήματα όπου παρουσιάζεται η αύξηση στο SIR εξόδο και για τις δύο πηγές s_1 και s_2 για 2,3,4 και 5 μικρόφωνα αντίστοιχα.



Σχήμα 6.2 Σχηματική αναπαράσταση της αύξησης του SIR της πηγής s_1 με την αύξηση του αριθμού των μικροφώνων.



Σχήμα 6.3 Σχηματική αναπαράσταση της αύξησης του SIR της πηγής s_2 με την αύξηση του αριθμού των μικροφώνων.

Τα πειραματικά αποτελέσματα δείχνουν ότι με αύξηση του αριθμού των μικροφώνων βελτιώνεται η απόδοση του αλγορίθμου. Βέβαια από έξι και πάνω μικρόφωνα το ποσοστό της βελτίωσης δεν είναι σημαντικό. Αυτό δικαιολογείται από τα γεγονότα ότι όση και να είναι η επιπρόσθετη πληροφορία από τα μικρόφωνα, αυτή δεν επηρεάζει άμεσα την πολυπλοκότητα του δικτύου διαχωρισμού, αφού οι βασικές δομικές μονάδες που το απαρτίζουν είναι απλά δίκτυα διαχωρισμού με δύο εισόδους και δύο εξόδους οι οποίες μπορούν και δουλεύουν αποδοτικά αφού η κάθε μία από αυτές χρειάζεται να υπολογίζει μόνο

τέσσερα φίλτρα διαχωρισμού. Τα βασικά δίκτυα διαχωρισμού μπορούν να επεξεργάζονται τα σήματα ομιλίας ανεξάρτητα (σε παράλληλη τοπολογία) μειώνοντας έτσι δραστικά τον χρόνο που απαιτείται για την επίλυση του υπέρ-ορισμένου διαχωρισμού.

6.4.1 Πραγματοποίηση μετρήσεων μεταβάλλοντας τις παραμέτρους του δωματίου

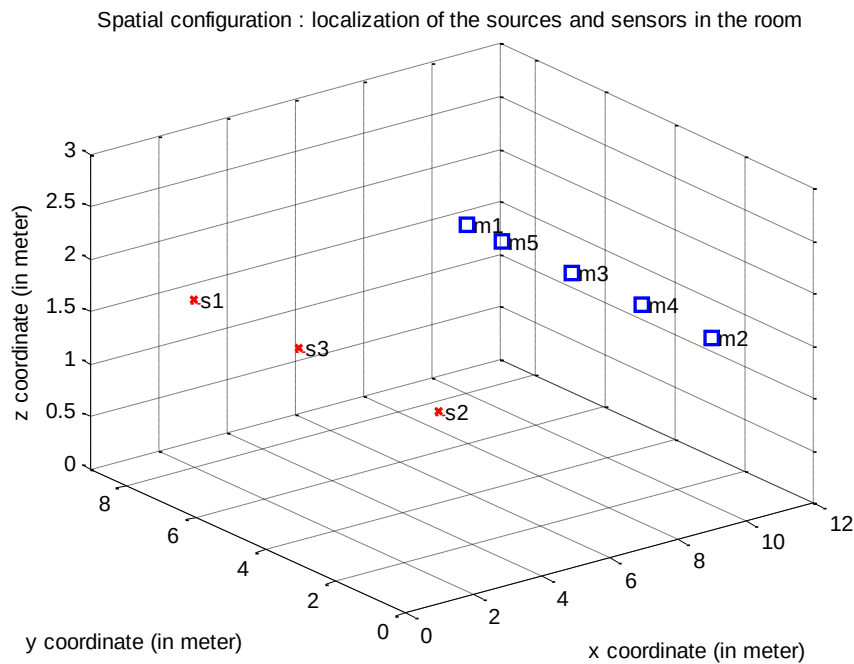
1^η Σειρά μετρήσεων:

Στην ενότητα αυτή θα μελετήσουμε αρχικά αν η μεταβολή του αριθμού των εικονικών πηγών μεταβάλλει το SIR των σημάτων. Οι παράμετροι που χρησιμοποιούμε στις μετρήσεις μας φαίνονται στον Πίνακα 6.1:

<i>Παράμετροι μετρήσεων</i>	
<i>Διάρκεια πηγών</i>	<i>5s</i>
<i>Αριθμός πηγών</i>	<i>3</i>
<i>Αριθμός μικροφώνων</i>	<i>5</i>
<i>Συντελεστής ανακλάσεως τοίχων</i>	<i>-0.2</i>
<i>Μήκος FFT</i>	<i>2048</i>
<i>Συντελεστής καθυστέρησης</i>	<i>0.75</i>
<i>Παράθυρο</i>	<i>Hanning</i>
<i>Αριθμός των FFT πλαισίων για μέση τιμή δείγματος</i>	<i>7</i>

Πίνακας 6.1 Παράμετροι μετρήσεων στην περίπτωση του υπερ-ορισμένου διαχωρισμού

Η απεικόνιση του δωματίου με $I = 3$ πηγές και $J = 5$ μικρόφωνα φαίνεται στο Σχήμα 6.4.



Σχήμα 6.4 Απεικόνιση των τριών πηγών και των πέντε μικροφώνων μέσα σε ένα δωμάτιο με ανακλάσεις.

Οι μετρήσεις που προκύπτουν μεταβάλλοντας τον αριθμό των εικονικών πηγών φαίνονται στον Πίνακα 6.2.

Αριθμός εικονικών πηγών	SIR			
	s_1	s_2	s_3	Συνολικό
70	21.8388	14.0243	9.38538	15.0828
40	21.8388	14.0243	9.38538	15.0828
10	21.8388	14.0243	9.38538	15.0828

Πίνακας 6.2 Μεταβολή του αριθμού των εικονικών πηγών και υπολογισμός SIR

Παρατηρούμε ότι το SIR εξόδου κάθε σήματος παραμένει σταθερό ανεξάρτητα από τον αριθμό των εικονικών πηγών που θεωρούμε.

2^η Σειρά μετρήσεων:

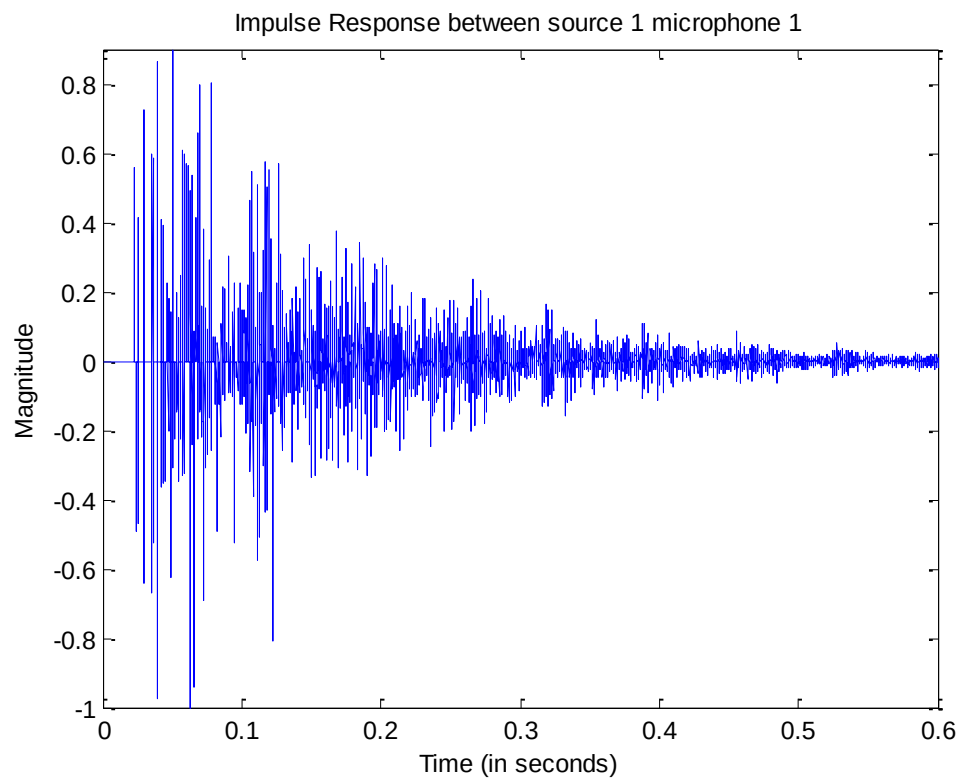
Μια δεύτερη σειρά μετρήσεων που πραγματοποιούμε είναι η μελέτη του αν και πόσο αλλάζει το SIR κάθε πηγής καθώς και το συνολικό SIR, αν μεταβάλλουμε τους συντελεστές ανακλάσεως των τοίχων του δωματίου. Οι τιμές που παίρνουν οι συντελεστές ανακλάσεως είναι μεταξύ -1 και 1. Παίρνουμε μετρήσεις για αρνητικό συντελεστή ισχύος με $\text{ref}=-0.9$ για μεγάλο T60, $\text{ref}=-0.2$

και $\text{ref}=-0.1$ για μικρό T60 καθώς και μετρήσεις για θετικό συντελεστή ισχύος με $\text{ref}=0.3$, $\text{ref}=0.5$ και $\text{ref}=0.7$. Οι μετρήσεις που προκύπτουν μεταβάλλοντας τους συντελεστές φαίνονται στον Πίνακα 6.3.

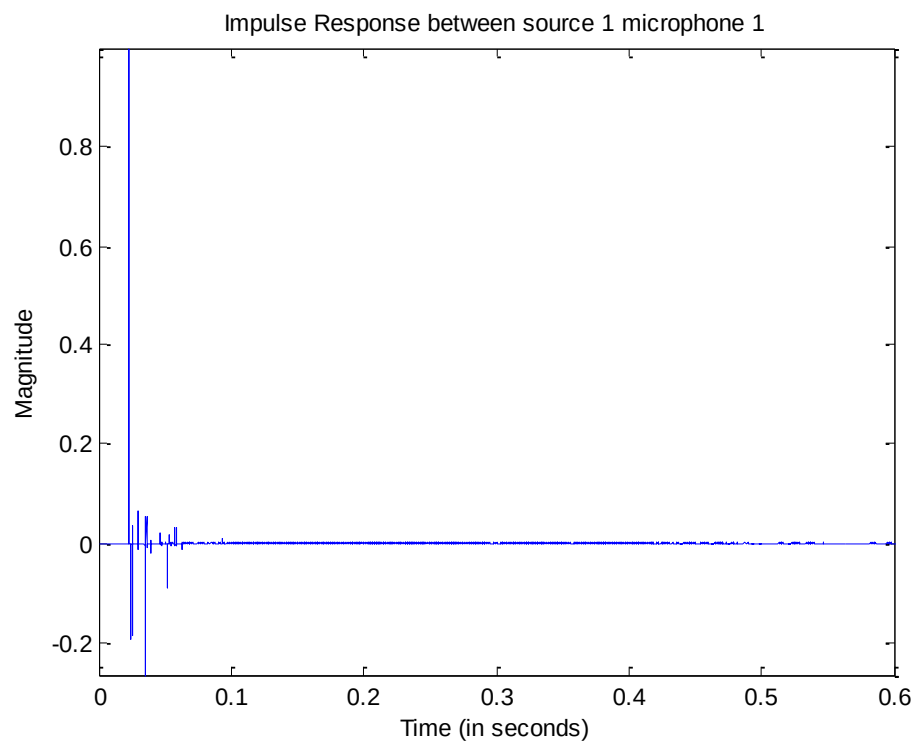
	SIR			
Συντελεστής ανακλάσεως τοίχων	s_1	s_2	s_3	Συνολικό
-0,9	5,5302	1,8233	3,9369	3,7635
-0,2	22,5643	25,7741	26,1759	24,8381
-0,1	24,6403	27,2118	25,4979	25,7833
0,3	21,7351	24,9142	22,3107	22,9867
0,5	14,6846	19,0723	15,577	16,4446
0,7	9,6811	11,0702	10,3382	10,3631

Πίνακας 6.3 Μεταβολή του συντελεστή ανακλάσεως και υπολογισμός SIR

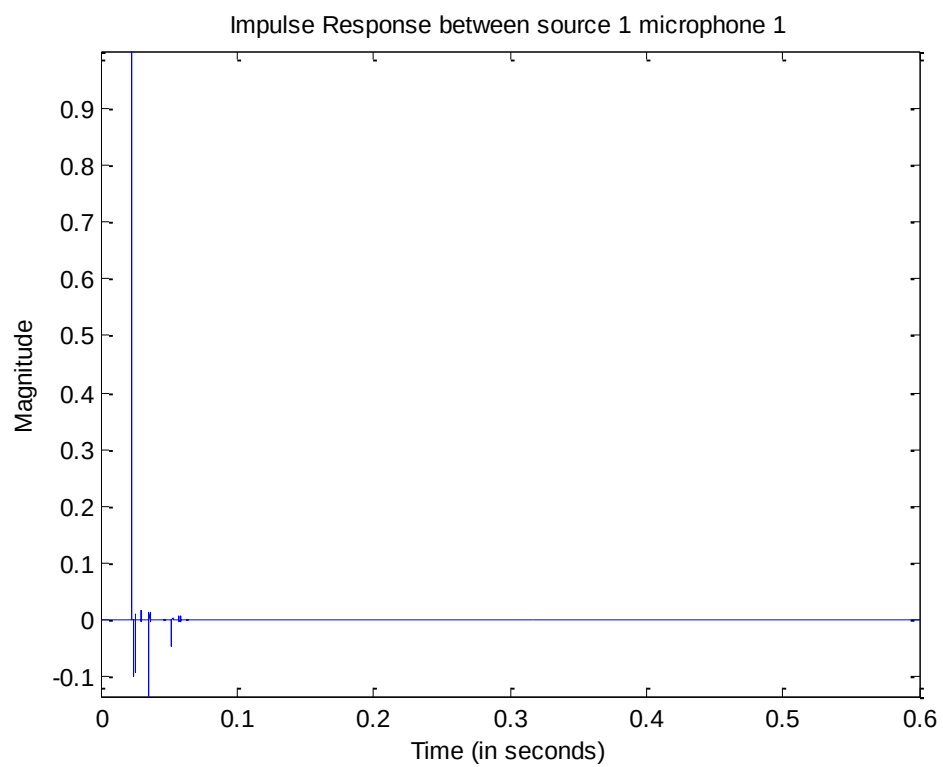
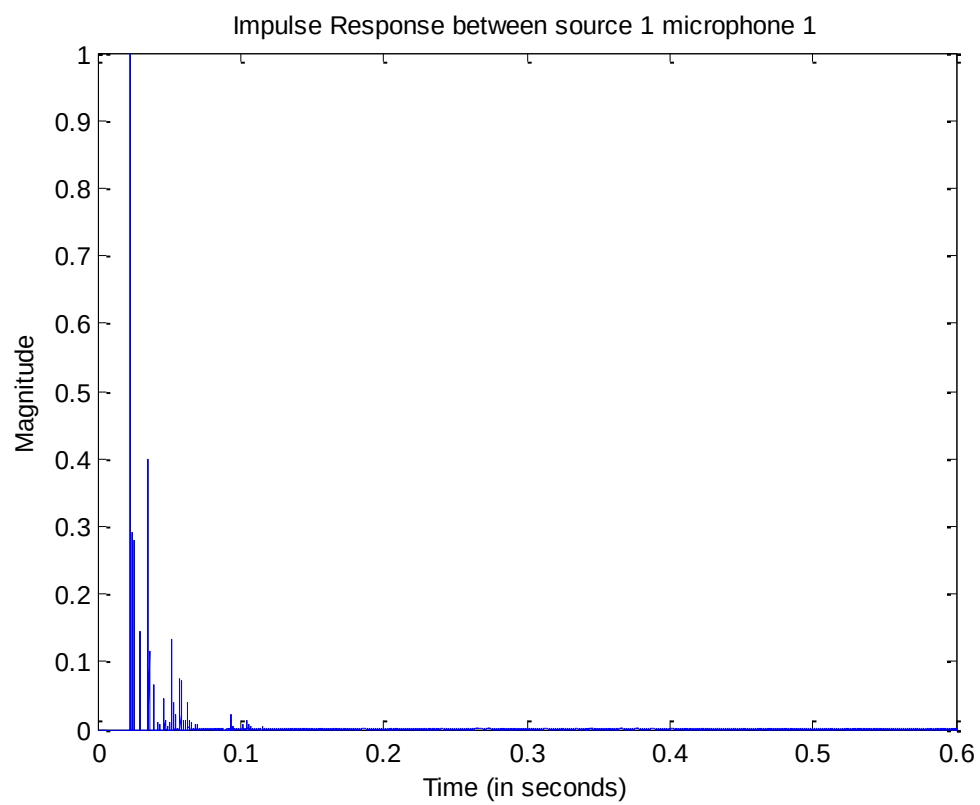
Στο Σχήμα 6.5 φαίνεται η κρουστική απόκριση μεταξύ της πηγής 1 και του μικροφώνου 1, για τις παραπάνω τιμές του συντελεστή ανακλάσεως (-0,9, -0,2, -0,1, 0,3, 0,5 και 0,7 αντίστοιχα):

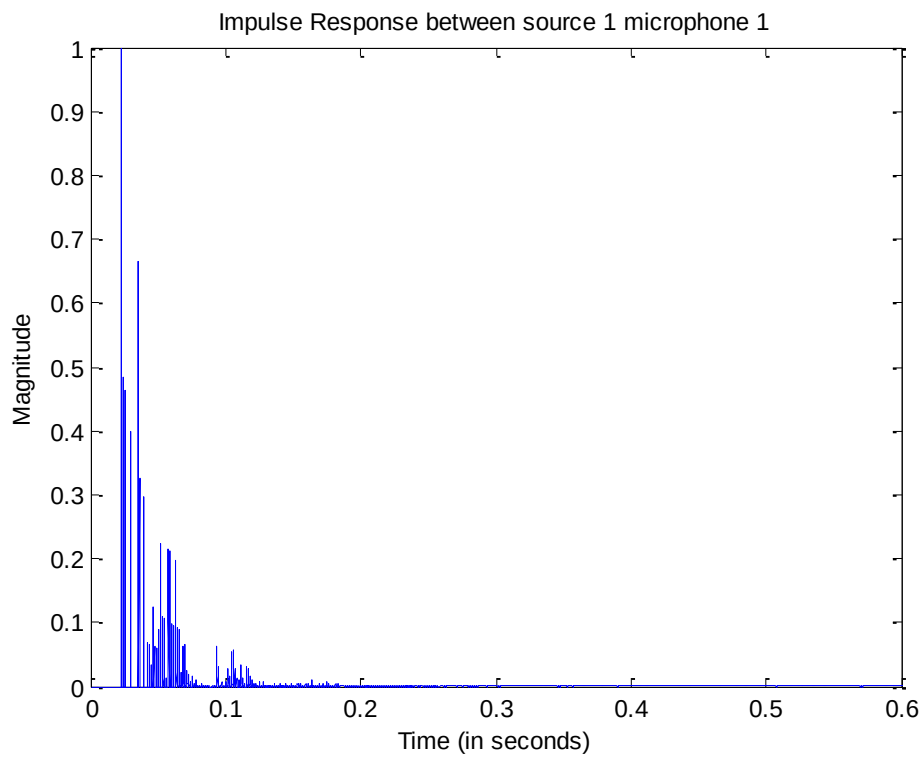


(α)

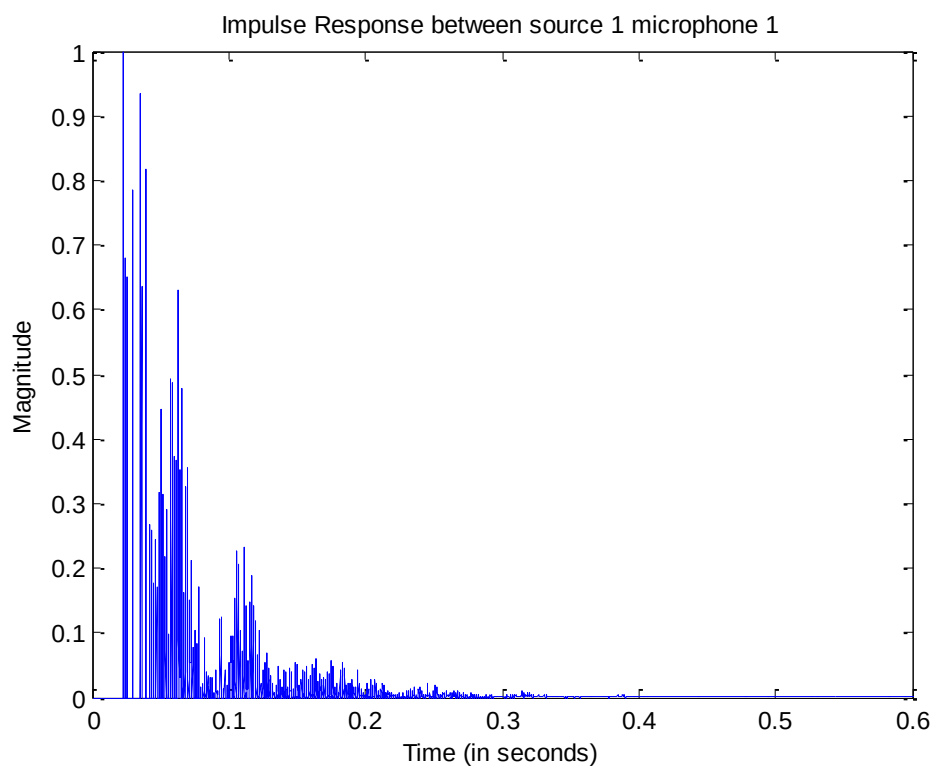


(β)

 (γ)  (δ)



(ε)



(σ)

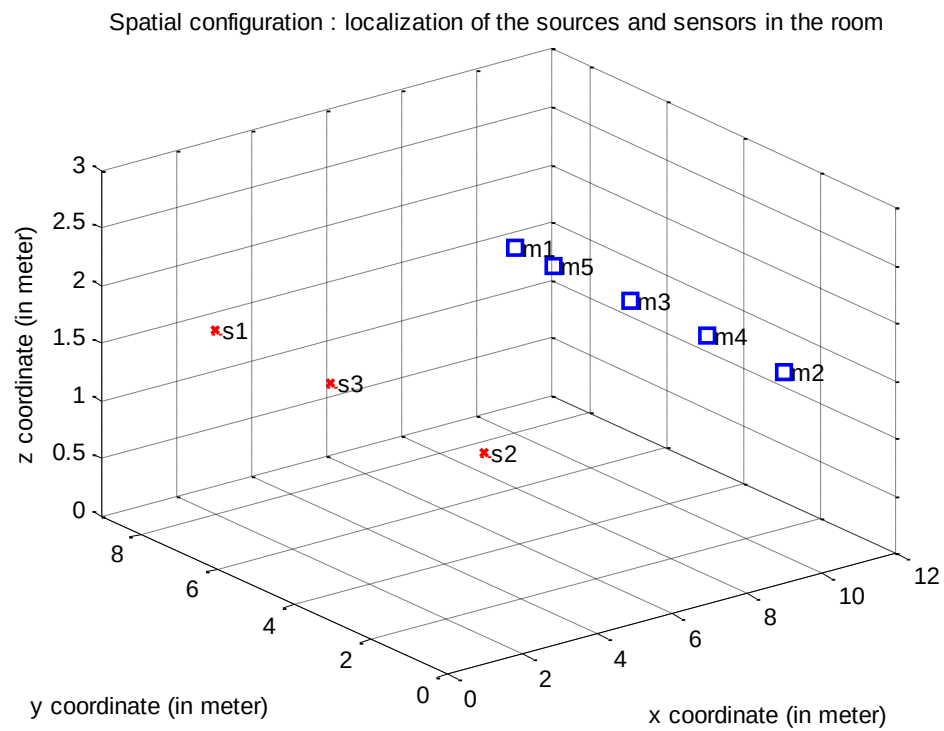
Σχήμα 6.5 Κρουστική απόκριση μεταξύ της πηγής 1 και του μικροφώνου 1 για συντελεστή ανακλάσεως (α) $ref=-0,9$, (β) $ref=-0,2$, (γ) $ref=-0,1$, (δ) $ref=0,3$, (ε) $ref=0,5$ και (στ) $ref=0,7$.

Παρατηρούμε λοιπόν, ότι όσο μεγαλύτερη θετική τιμή έχει ο συντελεστής ανακλάσεως των τοίχων τόσο μειώνεται η μεταβολή του σήματος προς παρεμβολή (SIR). Για αρνητικές τιμές του συντελεστή ανακλάσεως και συγκεκριμένα για συντελεστή ισχύος $ref=-0,1$ δηλαδή μικρό T60, έχουμε έναν πολύ καλό διαχωρισμό των σημάτων. Βέβαια η παραπάνω διαπίστωση ισχύει για τιμές μεταξύ $-0,9$ έως $-0,1$. Για τιμές $ref=-1$ και $ref=1$ δεν μπορεί να πραγματοποιηθεί ο διαχωρισμός των σημάτων.

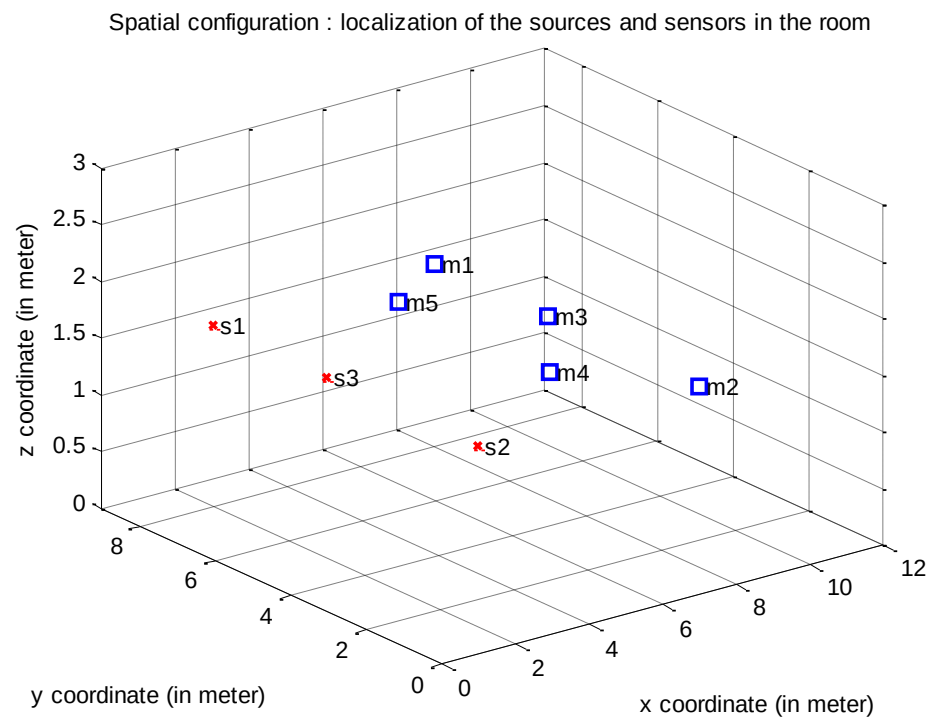
3^η Σειρά μετρήσεων

Τέλος, θα μελετήσουμε πώς επηρεάζει την ποιότητα του διαχωρισμού η μεταβολή της απόστασης μεταξύ των πηγών και των μικροφώνων. Οι μετρήσεις που παίρνουμε για συγκεκριμένες θέσεις των μικροφώνων στο δωμάτιο φαίνονται στον παρακάτω Πίνακα 6.4.

Οι απεικονίσεις του δωματίου για κάθε μία μέτρηση που πραγματοποιείται φαίνονται στο Σχήμα 6.6.

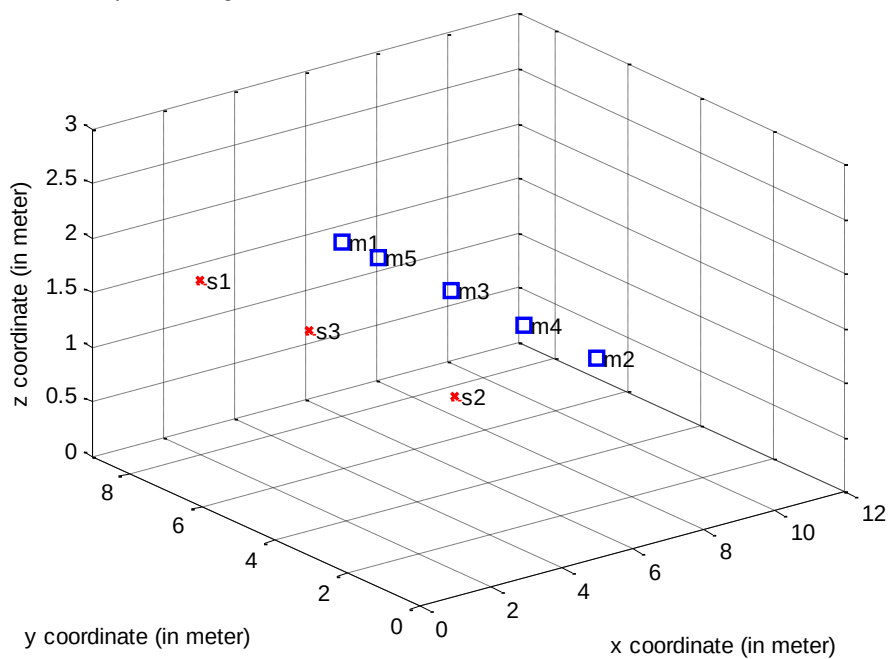


(α)



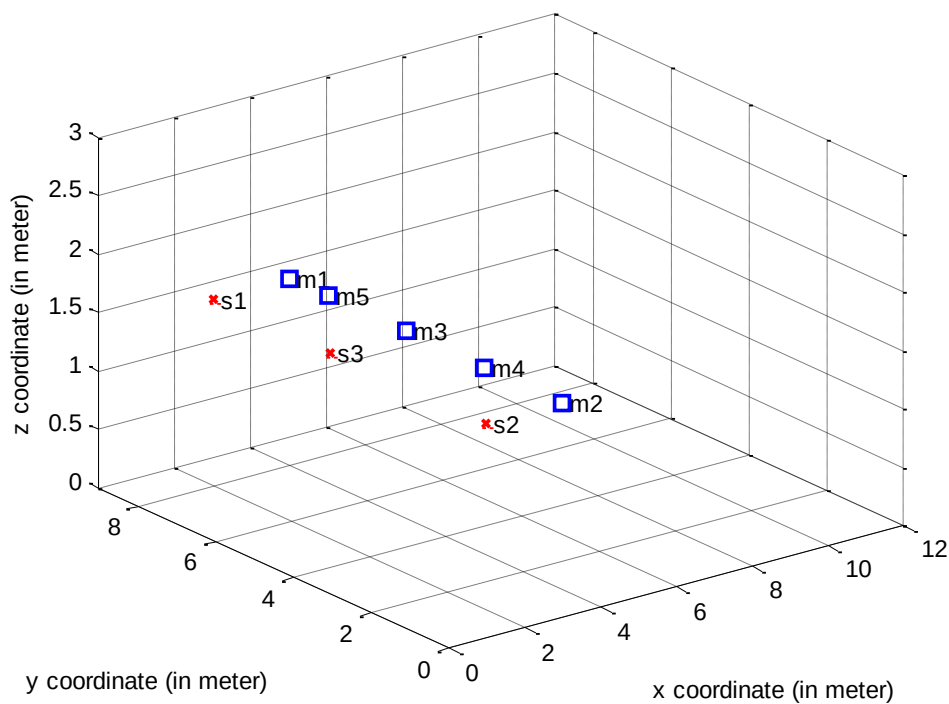
(β)

Spatial configuration : localization of the sources and sensors in the room



(γ)

Spatial configuration : localization of the sources and sensors in the room



(δ)

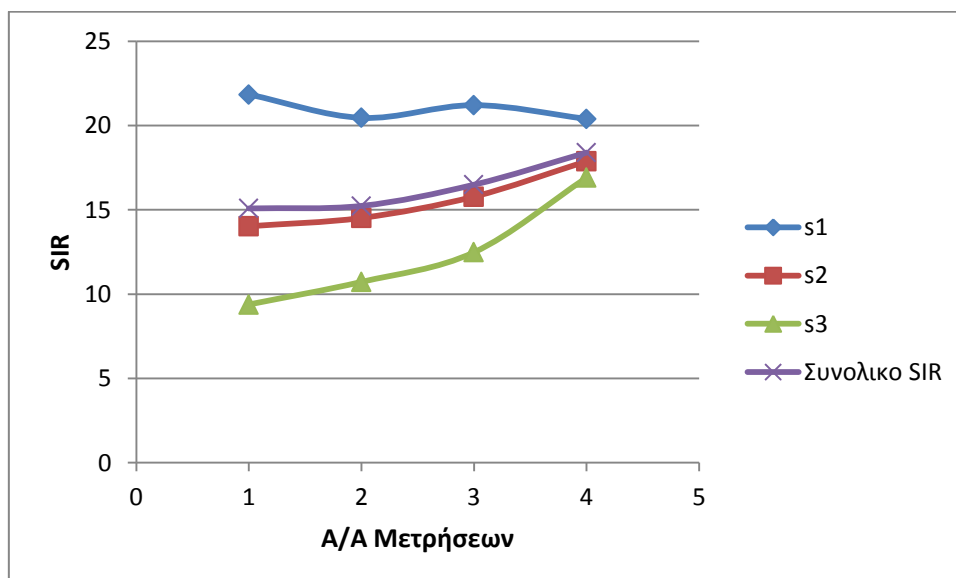
Σχήμα 6.6 Απεικόνιση της θέσης των πηγών και των μικροφώνων για κάθε μία μέτρηση (α) 1^η, (β) 2^η, (γ) 3^η και (δ) 4^η μέτρηση.

Οι τιμές του SIR που προκύπτουν και αντιστοιχούν σε κάθε μια από τις παραπάνω απεικονίσεις βρίσκονται στον Πίνακα 6.4.

Μετρήσεις	SIR			
	s_1	s_2	s_3	Συνολικό
1	21.8388	14.0243	9.38538	15.0828
2	20.4541	14.5176	10.7246	15.2321
3	21.2074	15.7816	12.4805	16.4898
4	20.3936	17.8792	16.9041	18.3923

Πίνακας 6.4 Μεταβολή των αποστάσεων μεταξύ πηγών και μικροφώνων-υπολογισμός SIR

Παρατηρούμε ότι η τιμή του SIR για τα s_2 και s_3 αυξάνεται όπως και η τιμή του συνολικού SIR. Αντίθετα το s_1 επηρεάζεται αρνητικά κατά τη μετακίνηση των μικροφώνων στη δεύτερη μέτρηση, θετικά στην τρίτη και αρνητικά στην τέταρτη. Αυτό πιθανόν συμβαίνει επειδή το s_1 βρίσκεται κοντά στον τοίχο οπότε οι ανακλάσεις είναι εντονότερες. Η γραφική απεικόνιση των μετρήσεων που προηγήθηκαν φαίνονται στο Σχήμα 6.7:



Σχήμα 6.7 Γραφική απεικόνιση της μεταβολής του SIR

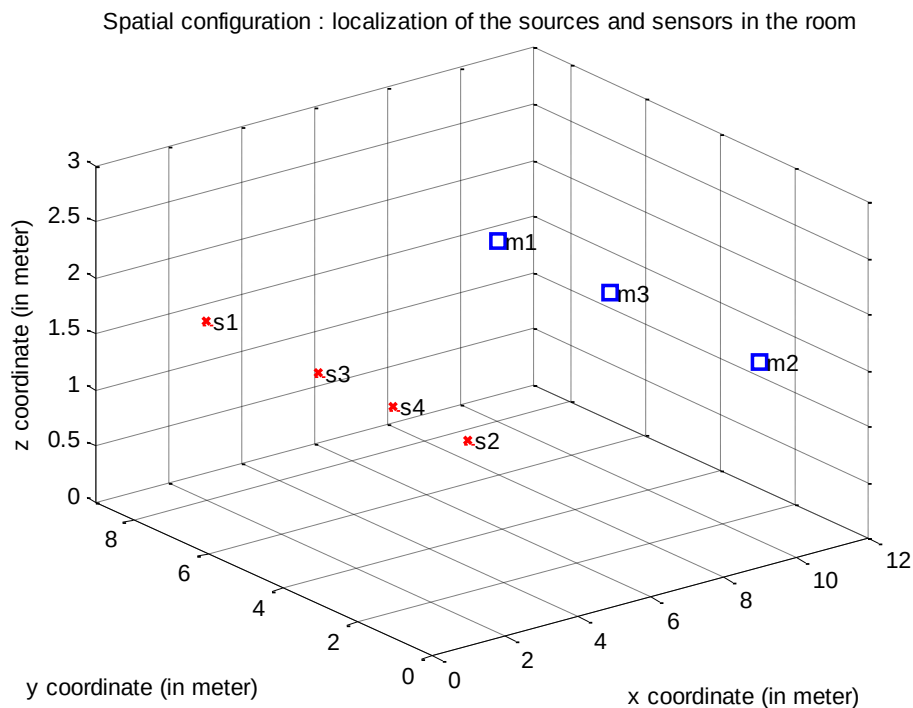
6.5 Υπο-ορισμένος διαχωρισμός σημάτων σε συνελικτική μίξη

Στην ενότητα αυτή θα μελετήσουμε τον υπερ-ορισμένο διαχωρισμό σημάτων σε συνελικτική χρησιμοποιώντας το μοντέλο PARAFAC για το διαχωρισμό. Αρχικά θα μελετήσουμε πώς επηρεάζεται η ποιότητα του διαχωρισμού όταν έχουμε μεταβολή των αποστάσεων των μικροφώνων σε σχέση με τις πηγές. Οι παράμετροι του συστήματός μας δίνονται στον Πίνακα 6.5.

Διάρκεια πηγών	20s
Αριθμός πηγών	4
Αριθμός μικροφώνων	3
Συντελεστής ανακλάσεως τοίχων	-0.2
Μήκος FFT	2048
Συντελεστής καθυστέρησης	0.75
Παράθυρο	Hanning
Αριθμός των FFT πλαισίων για μέση τιμή δείγματος	7

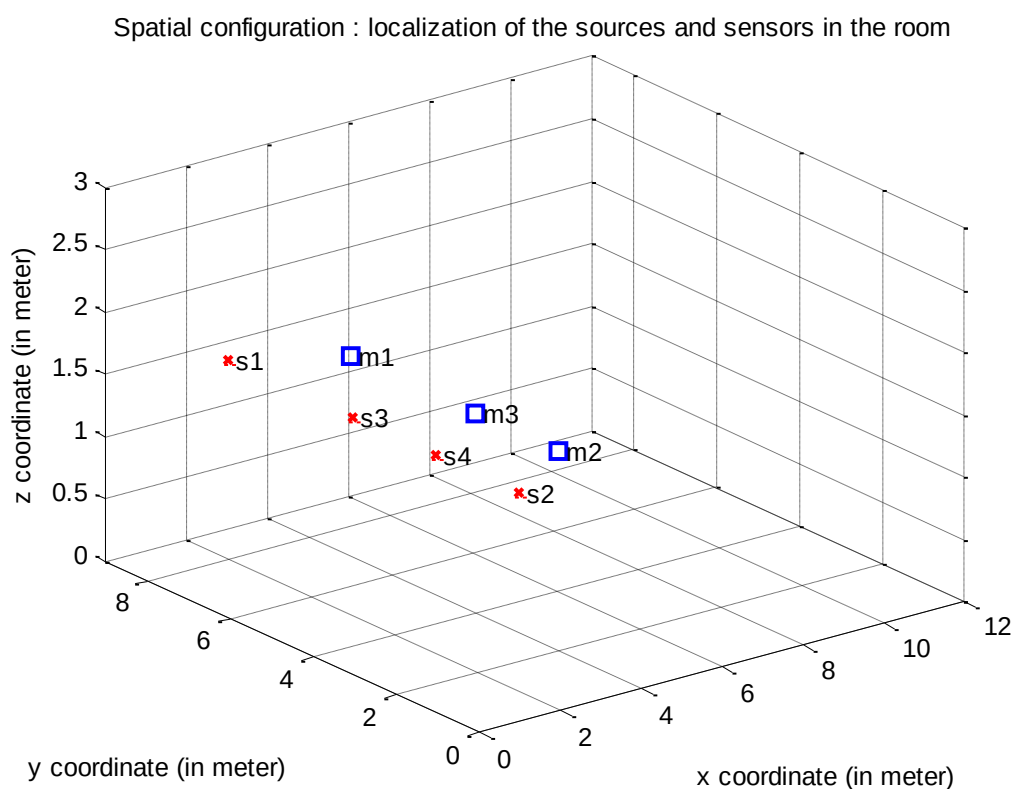
Πίνακας 6.5 Παράμετροι μετρήσεων στην περίπτωση του υπο-ορισμένου διαχωρισμού

Η απεικόνιση του δωματίου με $I = 4$ πηγές και $J = 3$ μικρόφωνα φαίνεται στο Σχήμα 6.8.



Σχήμα 6.8 Απεικόνιση των τεσσάρων πηγών και των τριών μικροφώνων μέσα σε ένα δωμάτιο με ανακλάσεις

Παρατηρούμε ότι μετά την εφαρμογή του μοντέλου PARAFAC δεν λαμβάνουμε ικανοποιητικά αποτελέσματα για το διαχωρισμό των σημάτων των πηγών. Ακόμα και όταν πλησιάζουμε τα μικρόφωνα πιο κοντά στους ομιλητές (Σχήμα 6.9) η απόδοση του συστήματος δε βελτιώνεται ικανοποιητικά. Τα διαχωρισμένα σήματα που προκύπτουν είναι σαν ακούμε τη μίξη όλων των σημάτων με τη διαφορά ότι κάθε φορά ένα έχει μεγαλύτερη ένταση.



Σχήμα 6.9 Μείωση της απόστασης μεταξύ των πηγών και των μικροφώνων.

Στη συνέχεια μελετάμε αν η μεταβολή του συντελεστή ανακλάσεως επηρεάζει την απόδοση. Συγκεκριμένα, παίρνουμε μετρήσεις για τιμές του συντελεστή ανακλάσεως $\text{ref} = -0,9, -0,2, -0,1$, στο δωμάτιο που φαίνεται στο Σχήμα 6.9. Στηριζόμενοι στα ακουστικά αποτελέσματα που προκύπτουν μετά την εφαρμογή του αλγορίθμου διαπιστώνουμε ότι όσο μειώνεται ο συντελεστής ανάκλασης βελτιώνεται η απόδοση του συστήματος χωρίς αυτό να σημαίνει ότι συντελείται επιτυχώς ο διαχωρισμός. Δηλαδή, μετά το διαχωρισμό κάθε σήμα πηγής έχει μεγαλύτερη ένταση από τα υπόλοιπα κάτι που δεν αποτελεί ικανοποιητική λύση.

ΠΑΡΑΡΤΗΜΑ

Ο αλγόριθμος fastICA που χρησιμοποιήθηκε για την υλοποίηση των μετρήσεων στο Κεφάλαιο 2 και στο Κεφάλαιο 4 είναι ο ακόλουθος:

```
function [Out1, Out2, Out3] = fastica(mixedsig, varargin)
%FASTICA - Fast Independent Component Analysis
%
% FastICA for Matlab 7.x and 6.x
% Version 2.5, October 19 2005
% Copyright (c) Hugo Gvert, Jarmo Hurri, Jaakko Srel, and Aapo
Hyvrinen.
%
% FASTICA(mixedsig) estimates the independent components from given
% multidimensional signals. Each row of matrix mixedsig is one
% observed signal. FASTICA uses Hyvarinen's fixed-point algorithm,
% see http://www.cis.hut.fi/projects/ica/fastica/. Output from the
% function depends on the number output arguments:
%
% [icasig] = FASTICA (mixedsig); the rows of icasig contain the
% estimated independent components.
%
% [icasig, A, W] = FASTICA (mixedsig); outputs the estimated
separating
% matrix W and the corresponding mixing matrix A.
%
% [A, W] = FASTICA (mixedsig); gives only the estimated mixing matrix
% A and the separating matrix W.
%
% Some optional arguments induce other output formats, see below.
%
% A graphical user interface for FASTICA can be launched by the
% command FASTICAG
%
% FASTICA can be called with numerous optional arguments. Optional
% arguments are given in parameter pairs, so that first argument is
% the name of the parameter and the next argument is the value for
% that parameter. Optional parameter pairs can be given in any order.
%
% OPTIONAL PARAMETERS:
%
% Parameter name      Values and description
%
%=====
%
% --Basic parameters in fixed-point algorithm:
%
% 'approach'          (string) The decorrelation approach used. Can
be
%                      symmetric ('symm'), i.e. estimate all the
%                      independent component in parallel, or
%                      deflation ('defl'), i.e. estimate independent
%                      component one-by-one like in projection pursuit.
%                      Default is 'defl'.
%
% 'numOfIC'           (integer) Number of independent components to
%                      be estimated. Default equals the dimension of data.
%
%=====
%
%=====
```

```

% --Choosing the nonlinearity:
%
% 'g'                (string) Chooses the nonlinearity g used in
%                    the fixed-point algorithm. Possible values:
%
%                    Value of 'g':      Nonlinearity used:
%                    'pow3' (default)    g(u)=u^3
%                    'tanh'              g(u)=tanh(a1*u)
%                    'gauss'             g(u)=u*exp(-a2*u^2/2)
%                    'skew'              g(u)=u^2
%
% 'finetune'          (string) Chooses the nonlinearity g used when
%                    fine-tuning. In addition to same values
%                    as for 'g', the possible value 'finetune' is:
%                    'off'               fine-tuning is disabled.
%
% 'a1'                (number) Parameter a1 used when g='tanh'.
%                    Default is 1.
% 'a2'                (number) Parameter a2 used when g='gaus'.
%                    Default is 1.
%
% 'mu'                (number) Step size. Default is 1.
%                    If the value of mu is other than 1, then the
%                    program will use the stabilized version of
the
%                    algorithm (see also parameter
'stabilization').
%
%
% 'stabilization'      (string) Values 'on' or 'off'. Default 'off'.
% This parameter controls whether the program uses the stabilized
% version of the algorithm or not. If the stabilization is on, then
% the value of mu can momentarily be halved if the program senses that
% the algorithm is stuck between two points (this is called a stroke).
% Also if there
%
%                    is no convergence before half of the maximum
%                    number of iterations has been reached then mu
%                    will be halved for the rest of the rounds.
%
%=====
==
% --Controlling convergence:
%
% 'epsilon'            (number) Stopping criterion. Default is
0.0001.
%
% 'maxNumIterations'    (integer) Maximum number of iterations.
%                    Default is 1000.
%
% 'maxFinetune'         (integer) Maximum number of iterations in
%                    fine-tuning. Default 100.
%
% 'sampleSize'         (number) [0 - 1] Percentage of samples used in
%                    one iteration. Samples are chosen in random.
%                    Default is 1 (all samples).
%
% 'initGuess'          (matrix) Initial guess for A. Default is random.
%                    You can now do a "one more" like this:
%                    [ica, A, W] = fastica(mix, 'numOfIC',3);
%                    [ica2, A2, W2] = fastica(mix, 'initGuess', A, 'numOfIC', 4);
%

```

```

%=====
==
% --Graphics and text output:
%
% 'verbose'          (string) Either 'on' or 'off'. Default is
%                    'on': report progress of algorithm in text
format.
%
% 'displayMode'      (string) Plot running estimates of
independent
%                    components: 'signals', 'basis', 'filters' or
%                    'off'. Default is 'off'.
%
% 'displayInterval'  Number of iterations between plots.
%                    Default is 1 (plot after every iteration).
%
%=====
==
% --Controlling reduction of dimension and whitening:
%
% Reduction of dimension is controlled by 'firstEig' and 'lastEig',
or
% alternatively by 'interactivePCA'.
%
% 'firstEig'          (integer) This and 'lastEig' specify the
range for
%                    eigenvalues that are retained, 'firstEig' is
%                    the index of largest eigenvalue to be
%                    retained. Default is 1.
%
% 'lastEig'           (integer) This is the index of the last
(smallest)
%                    eigenvalue to be retained. Default equals the
%                    dimension of data.
%
% 'interactivePCA'    (string) Either 'on' or 'off'. When set 'on',
the
%                    eigenvalues are shown to the user and the
%                    range can be specified interactively. Default
%                    is 'off'. Can also be set to 'gui'. Then the
user
%                    can use the same GUI that's in FASTICAG.
%
% If you already know the eigenvalue decomposition of the covariance
% matrix, you can avoid computing it again by giving it with the
% following options:
%
% 'pcaE'              (matrix) Eigenvectors
% 'pcaD'              (matrix) Eigenvalues
%
% If you already know the whitened data, you can give it directly to
% the algorithm using the following options:
%
% 'whiteSig'          (matrix) Whitened signal
% 'whiteMat'          (matrix) Whitening matrix
% 'dewhiteMat'        (matrix) dewhitening matrix
%
% If values for all the 'whiteSig', 'whiteSig' and 'dewhiteMat' are
% supplied, they will be used in computing the ICA. PCA and whitening
% are not performed. Though 'mixedsig' is not used in the main
% algorithm it still must be entered - some values are still

```

```

% calculated from it.
%
% Performing preprocessing only is possible by the option:
%
% 'only'          (string) Compute only PCA i.e. reduction of
%                  dimension ('pca') or only PCA plus whitening
%                  ('white'). Default is 'all': do ICA estimation
%                  as well. This option changes the output
%                  format accordingly. For example:
%
%[whitesig, WM, DWM] = FASTICA(mixedsig, 'only', 'white')
%                  returns the whitened signals, the whitening
%matrix            (WM) and the dewhitening matrix (DWM).
%                  (See %alsoWHITENV.) In FastICA the whitening
%matrix performs whitening and the reduction of dimension.
%                  %Dewhitening matrix is the pseudoinverse of
%whitening matrix.
%
%
%                  [E, D] = FASTICA(mixedsig, 'only', 'pca')
%                  returns the eigenvector (E) and diagonal
%                  eigenvalue (D) matrices containing the
%                  selected subspaces.
%
%=====
% EXAMPLES
%
%[icasig] = FASTICA (mixedsig, 'approach', 'symm', 'g', 'tanh');
%          Do ICA with tanh nonlinearity and in parallel (like
%          maximum likelihood estimation for supergaussian data).
%
%          [icasig] = FASTICA (mixedsig, 'lastEig', 10, 'numOfIC', 3);
%          Reduce dimension to 10, and estimate only 3
%          independent components.
%
%[icasig]=FASTICA (mixedsig, 'verbose', 'off', 'displayMode', 'off');
%          Don't output convergence reports and don't plot
%          independent components.
%
%
% A graphical user interface for FASTICA can be launched by the
% command FASTICAG
%
% See also FASTICAG

% @(#)$Id: fastica.m,v 1.14 2005/10/19 13:05:34 jarmo Exp $

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Check some basic requirements of the data
if nargin == 0,
    error ('You must supply the mixed data as input argument.');
```

```

end

if length (size (mixedsig)) > 2,
    error ('Input data can not have more than two dimensions.');
```

```

end

if any (any (isnan (mixedsig))),
    error ('Input data contains NaN's.');
```

```

end

if ~isa (mixedsig, 'double')
    fprintf ('Warning: converting input data into regular (double)
precision.\n');
    mixedsig = double (mixedsig);
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Remove the mean and check the data

[mixedsig, mixedmean] = remmean(mixedsig);

[Dim, NumOfSampl] = size(mixedsig);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Default values for optional parameters

% All
verbose          = 'on';

% Default values for 'pcamat' parameters
firstEig         = 1;
lastEig          = Dim;
interactivePCA    = 'off';

% Default values for 'fpica' parameters
approach         = 'defl';
numOfIC          = Dim;
g                = 'pow3';
finetune         = 'off';
a1               = 1;
a2               = 1;
myy              = 1;
stabilization    = 'off';
epsilon          = 0.0001;
maxNumIterations = 1000;
maxFinetune      = 5;
initState        = 'rand';
guess            = 0;
sampleSize       = 1;
displayMode      = 'off';
displayInterval  = 1;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Parameters for fastICA - i.e. this file

b_verbose = 1;
jumpPCA = 0;
jumpWhitening = 0;
only = 3;
userNumOfIC = 0;

```

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Read the optional parameters

if (rem(length(varargin),2)==1)
    error('Optional parameters should always go by pairs');
else
    for i=1:2:(length(varargin)-1)
        if ~ischar (varargin{i}),
            error (['Unknown type of optional parameter name (parameter'
...
                ' names must be strings).']);
        end
        % change the value of parameter
        switch lower (varargin{i})
            case 'stabilization'
                stabilization = lower (varargin{i+1});
            case 'maxfinetune'
                maxFinetune = varargin{i+1};
            case 'samplesize'
                sampleSize = varargin{i+1};
            case 'verbose'
                verbose = lower (varargin{i+1});
                % silence this program also
                if strcmp (verbose, 'off'), b_verbose = 0; end
            case 'firsteig'
                firstEig = varargin{i+1};
            case 'lasteig'
                lastEig = varargin{i+1};
            case 'interactivepca'
                interactivePCA = lower (varargin{i+1});
            case 'approach'
                approach = lower (varargin{i+1});
            case 'numofic'
                numOfIC = varargin{i+1};
                % User has supplied new value for numOfIC.
                % We'll use this information later on...
                userNumOfIC = 1;
            case 'g'
                g = lower (varargin{i+1});
            case 'finetune'
                finetune = lower (varargin{i+1});
            case 'a1'
                a1 = varargin{i+1};
            case 'a2'
                a2 = varargin{i+1};
            case {'mu', 'myy'}
                myy = varargin{i+1};
            case 'epsilon'
                epsilon = varargin{i+1};
            case 'maxnumiterations'
                maxNumIterations = varargin{i+1};
            case 'initguess'
                % no use setting 'guess' if the 'initState' is not set
                initState = 'guess';
                guess = varargin{i+1};
            case 'displaymode'
                displayMode = lower (varargin{i+1});
            case 'displayinterval'
                displayInterval = varargin{i+1};
            case 'pcae'

```

```

        % calculate if there are enough parameters to skip PCA
        jumpPCA = jumpPCA + 1;
        E = varargin{i+1};
    case 'pcad'
        % calculate if there are enough parameters to skip PCA
        jumpPCA = jumpPCA + 1;
        D = varargin{i+1};
    case 'whitesig'
        % calculate if there are enough parameters to skip PCA and
whitening
        jumpWhitening = jumpWhitening + 1;
        whitesig = varargin{i+1};
    case 'whitemat'
        % calculate if there are enough parameters to skip PCA and
whitening
        jumpWhitening = jumpWhitening + 1;
        whiteningMatrix = varargin{i+1};
    case 'dewhitemat'
        % calculate if there are enough parameters to skip PCA and
whitening
        jumpWhitening = jumpWhitening + 1;
        dewhitenMatrix = varargin{i+1};
    case 'only'
        % if the user only wants to calculate PCA or...
        switch lower (varargin{i+1})
            case 'pca'
                only = 1;
            case 'white'
                only = 2;
            case 'all'
                only = 3;
        end
    otherwise
        % Hmmm, something wrong with the parameter string
        error(['Unrecognized parameter: ' varargin{i} '']);
    end;
end;
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% print information about data
if b_verbose
    fprintf('Number of signals: %d\n', Dim);
    fprintf('Number of samples: %d\n', NumOfSampl);
end

% Check if the data has been entered the wrong way,
% but warn only... it may be on purpose

if Dim > NumOfSampl
    if b_verbose
        fprintf('Warning: ');
        fprintf('The signal matrix may be oriented in the wrong way.\n');
        fprintf('In that case transpose the matrix.\n\n');
    end
end
end

```

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Calculating PCA

% We need the results of PCA for whitening, but if we don't
% need to do whitening... then we dont need PCA...
if jumpWhitening == 3
    if b_verbose,
        fprintf ('Whitened signal and corresponding matrises
supplied.\n');
        fprintf ('PCA calculations not needed.\n');
    end;
else

    % OK, so first we need to calculate PCA
    % Check to see if we already have the PCA data
    if jumpPCA == 2,
        if b_verbose,
            fprintf ('Values for PCA calculations supplied.\n');
            fprintf ('PCA calculations not needed.\n');
        end;
    else
        % display notice if the user entered one, but not both, of E and
        D.
        if (jumpPCA > 0) & (b_verbose),
            fprintf ('You must suply all of these in order to jump
PCA:\n');
            fprintf ('''pcaE'', 'pcaD''.\n');
        end;

        % Calculate PCA
        [E, D]=pcamat(mixedsig, firstEig, lastEig, interactivePCA,
verbose);
    end
end

% skip the rest if user only wanted PCA
if only > 1

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
    % Whitening the data

    % Check to see if the whitening is needed...
    if jumpWhitening == 3,
        if b_verbose,
            fprintf ('Whitening not needed.\n');
        end;
    else

        % Whitening is needed
        % display notice if the user entered some of the whitening info,
        %but not all.
        if (jumpWhitening > 0) & (b_verbose),
            fprintf ('You must suply all of these in order to jump
whitening:\n');
            fprintf ('''whiteSig'', 'whiteMat'', 'dewhiteMat''.\n');
        end;
    end;
end

```

```

        % Calculate the whitening
        [whitesig, whiteningMatrix, dewateringMatrix] = whitenv ...
            (mixedsig, E, D, verbose);
    end

end % if only > 1

% skip the rest if user only wanted PCA and whitening
if only > 2

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
    % Calculating the ICA

    % Check some parameters
    % The dimension of the data may have been reduced during PCA
    %calculations.
    % The original dimension is calculated from the data by default,
    %and the
    % number of IC is by default set to equal that dimension.

    Dim = size(whitesig, 1);

    % The number of IC's must be less or equal to the dimension of data
    if numOfIC > Dim
        numOfIC = Dim;
        % Show warning only if verbose = 'on' and user supplied a value
    for 'numOfIC'
        if (b_verbose & userNumOfIC)
            fprintf('Warning: estimating only %d independent components\n',
numOfIC);
            fprintf('(Can''t estimate more independent components than
dimension of data)\n');
        end
    end

    % Calculate the ICA with fixed point algorithm.
    [A, W] = fpica (whitesig, whiteningMatrix, dewateringMatrix,
approach, ...
        numOfIC, g, finetune, a1, a2, myy, stabilization, epsilon,
...
        maxNumIterations, maxFinetune, initState, guess,
sampleSize, ...
        displayMode, displayInterval, verbose);

    % Check for valid return
    if ~isempty(W)
        % Add the mean back in.
        if b_verbose
            fprintf('Adding the mean back to the data.\n');
        end
        icasig = W * mixedsig + (W * mixedmean) * ones(1, NumOfSampl);
        %icasig = W * mixedsig;
        if b_verbose & ...
            (max(abs(W * mixedmean)) > 1e-9) & ...
            (strcmp(displayMode,'signals') | strcmp(displayMode,'on'))
            fprintf('Note that the plots don''t have the mean added.\n');
        end
    else

```

```

        icasig = [];
    end

end % if only > 2

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% The output depends on the number of output parameters
% and the 'only' parameter.

if only == 1      % only PCA
    Out1 = E;
    Out2 = D;
elseif only == 2 % only PCA & whitening
    if nargout == 2
        Out1 = whiteningMatrix;
        Out2 = dwhiteningMatrix;
    else
        Out1 = whitesig;
        Out2 = whiteningMatrix;
        Out3 = dwhiteningMatrix;
    end
else            % ICA
    if nargout == 2
        Out1 = A;
        Out2 = W;
    else
        Out1 = icasig;
        Out2 = A;
        Out3 = W;
    end
end
end

```

Με βάση τον παραπάνω αλγόριθμο fastICA οι λειτουργίες που συντελούνται για τον διαχωρισμό των σημάτων είναι οι εξής:

```

Number of signals: 2
Number of samples: 163830
Calculating covariance...
Dimension not reduced.
Selected [ 2 ] dimensions.
Smallest remaining (non-zero) eigenvalue [ 0.00210821 ]
Largest remaining (non-zero) eigenvalue [ 0.0445168 ]
Sum of removed eigenvalues [ 0 ]
[ 100 ] % of (non-zero) eigenvalues retained.
Whitening...
Check: covariance differs from identity by [ 8.61533e-014 ].
Used approach [ defl ].
Used nonlinearity [ pow3 ].
Starting ICA calculation...
IC 1 .....computed ( 8 steps )
IC 2 ..computed ( 2 steps )

```

Done.

Adding the mean back to the data.

Στο Κεφάλαιο 6 οι μετρήσεις πραγματοποιήθηκαν με βάση τον ακόλουθο αλγόριθμο:

```
%DEMO script for the Blind Speech Separation of convolutive mixtures
%based on CANDECOMP/PARAFAC, see paper D. Nion, K. N. Mokios, N. D.
%Sidiropoulos and A. Potamianos, "Batch and Adaptive PARAFAC-based
%Blind Separation of Convolutive Speech Mixtures", IEEE Trans. on
%Audio, Speech and Language Processing, Vol. 18, No. 6, pp. 1193-
%1207, August 2010.
```

```
% DEMO for a 3 by 5 over-determined case
```

```
clear all
close all
clc
```

```
%*****
%*****
```

```
%----- Create Mixtures -----
%-----
```

```
% Open load_data.m in build_mixtures folder to change the following
parameters:
```

```
% - the choice of the sources that are mixed
% - the room dimensions
% - the reverberation coefficients of the walls
% - the locations of sources
% - the locations of microphones
```

```
%*****
%*****
```

```
cd build_mixtures
```

```
%I=input('Number of sources? ');
```

```
I=2;
```

```
%J=input('Number of microphones? ');
```

```
J=7;
```

```
%ts=input('Choose duration of recording between 1 and 25 seconds ');
```

```
ts=5;
```

```
mes = waitbar(1, 'Building the mixtures');
```

```
load_data;
```

```
close (mes)
```

```
% Check that uniqueness condition is satisfied (see table in the
paper)
```

```
flagJ=0;
```

```
if J>=4;flagJ=1;end
```

```
if I*(I-1)/2 > (J*(J-1)/4)*(J*(J-1)/2+1)-(factorial(J)/(factorial(J-
1)*factorial(4)))*flagJ
```

```
    error('Uniqueness is not guaranteed for this choice of I and J')
```

```
end
```

```
%*****
%*****
```

```
%----- Parameters for the separation -----
%-----
```

```

%*****
%*****
% Parameters useful for computation of the second-order statistics
T = 2048; % Number of frequency bins (length of FFT)
ovlp = 0.75; % Percentage of overlap between two
consecutive FFT frames
wind = 'hanning'; % Window for FFT computation (e.g.
'hanning','hamming','kaiser',etc)
navg = 7; % Number of consecutive overlapping FFT frames
that are used to compute the
% sample mean estimate of covariance matrices

% Parameters of the PARAFAC-based separation stage
tol = 1e-8; % Tolerance to stop PARAFAC iterations
Niter = 500; % Maximum number of PARAFAC iterations

% Parameters of the sources spectra alignment stage (i.e., solving
%the permutation problem)
perm_strategy='power'; % string ('dominance' or 'power') to choose
%the kind of data that are clustered in the permutation solver
perm_method='kmeans'; % string ('iterative' or 'kmeans') to
choose the way the power porfiles are clustered
perm_deflation=0; % 0 or 1. If 0, then I! permutations are
%tried at each frequency.
% If 1, then the permutation is found by a deflation procedure.

%*****
%*****
%----- Display Summary of parameters -----
%-----
%*****
%*****
disp('-----')
disp('-----')
disp('----- Summary of Parameters -----')
disp('-----')
disp('-----')
disp(['Source duration (in secs.)',num2str(ts)]);
disp(['Number of sources:',int2str(I)]);
disp(['Number of microphones:',int2str(J)]);
disp(['Reflection coefficient of the walls:',num2str(ref)]);
disp(['FFT length:',num2str(T)]);
disp(['Overlap coefficient:',num2str(ovlp)]);
disp(['Window:',wind]);
disp(['Number of FFT frames for sample mean:',num2str(navg)]);
disp('-----')
disp(' ')
disp(' ')
%----- Compute Input SIR -----
disp('-----')
disp('----- Input SIR (# of microphones x # of sources) -----')
disp('-----')
disp('-----')
[SIR_input]=Input_SIR(convolved)

```

```

disp('-----')
disp(' ')
disp(' ')

cd ..
%*****
%----- START SEPARATION-----
%*****

[y,Wt,Wf,Hf]=sobi_conv_cp3(x,I,T,ovlp,wind,navg,tol,Niter,perm_strategy,perm_method,perm_deflation);
% [y,Wt,Wf,Hf]=sobi_conv_cp3(x,I); %--> to use all default parameters

if J>=I;
    [SIR,SIR_source]=output_SIR(Wt,convolved);
    disp('-----')
    disp('----- Output SIR -----')
    disp('-----')
    disp('-----')
    disp(['Output SIR averaged over all sources = ',num2str(SIR)])
    disp(['Output SIR for each source = ',num2str(SIR_source)])
    disp(' ')
end

disp('-----')
disp(' ')
disp('----- Listenning of the two first mixtures -----')
disp('-----')
disp('-----')
for mi=1:2
    disp(['Strike any key to listen to mixture at sensor ',num2str(mi)])
    pause
    sound(20*x(:,mi)/norm(x(:,mi)),Fs);
end
disp(' ')

disp('-----')
disp(' ')
disp('----- Listenning of the separated sources -----')
disp('-----')
disp('-----')
for so=1:I
    disp(['Strike any key to listen to the separated source ',num2str(so)])
    pause
    sound(20*y(:,so)/norm(y(:,so)),Fs);
end

```

Αλγόριθμος υπολογισμού SIR για την περίπτωση της συνελκτικής μίξης:

```

function [SIR_mean,SIR]=performance(Wt,convolved,bL)

%-----
%-----
% Computes the SIR to evaluate the separation performance, in over-
determined cases
% This function uses recordings of each microphone when only one
source
% at a time is active. In practice we don't have access to these
recordings
% but in the demo we do so we use them only for purpose of
performance
% evaluation.
% See equation (3) of paper
% "Evaluation of Blind Signal Separation Methods"
% by D. Schobben, K. Torkkola and P. Smaragdis
%-----
---
% INPUTS:
% - convolved: N x J x I (N samples, J mic, I sources)
%   holds the recordings when only one source at a time is active
% - Wt: I x J x T (I sources, J mic, T freq bins )
%   demixing filters in the time domain.
%-----
-----
% OUTPUTS:
% - SIR_mean: SIR in dB averaged over all sources
% - SIR: vector of I SIR values, one for each source
%-----
-----

N=size(convolved,1);
J=size(convolved,2);
I=size(convolved,3);

for i=1:I
    ys(:, :, i) = demix(Wt,convolved(:, :, i)); % of size NxIxI
end
power_all=squeeze(sum(ys.^2)); % IxI, for a given row holds the
power of target source and powers of cross-talks

for i=1:I
    id_all=[1:I];
    id_target=find(power_all(i, :)==max(power_all(i, :))); % to
identify target signal (max power)
    power_target=power_all(i,id_target);
    id_jammer=id_all; id_jammer(find(id_jammer==id_target))=[];
    power_jammer=sum(sum(ys(:, i, id_jammer).^2),3); % sum of powers
of cross-talk signals
    SIR(i)=10*log10(power_target/power_jammer);
end

SIR_mean=mean(SIR);

%*****
%*****
function y = demix(Wt, x)

[I J T] = size(Wt);
[N J] = size(x);

```

```

% allocate memory for y
y = zeros(N, I);

% do the convolution using overlap-add method
for i = 1:I
    for j = 1:J
        y(:, i) = y(:, i) + fftfilt(Wt(i, j, :), x(:, j));
    end
end

```


ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] E. Magno. La coarticolazione, Introduzione agli aspetti dinamici della produzione della parola. *CLESP*, Padova, 1980.
- [2] E. Farnetani. Coarticulation and connected speech processes. *Handbook of Phonetic Sciences*, W. Hardcastle and J. Laver, Blackwell, pp. 371-404, 1997.
- [3] M. Eskenazy. Trends in speaking style research. *Proc. EuroSpeech*, pp. 501-509, 1993.
- [4] J. Llisterri. Speaking styles in speech research. *Elsnet ESCA salt Workshop on Integrating Speech and Natural Language*, pp. 17-37, 1992.
- [5] J. Laver. The phonetic description of voice quality, *CUP*, 1980.
- [6] T. Shipp, H. Hollien. Perception of the ageing male voice. *Journal of Speech and hearing research*, 12, pp. 703-710, 1969.
- [7] E. Lombard. Le signe de l'elevation de la voix. *Annales Maladies de l'Oreille, Latynx, Nez et Pharynx*, 37, pp. 101-119, 1911.
- [8] D. Klatt, L. Klatt. Analysis synthesis and perception of voice quality variations among female and male talkers, *JASA*, 87(2): 820-857, 1990.
- [9] M. Green, L. Matheson. *The voice and its disorders*. Whurr Publishers, 1989.
- [10] W. Labov. Sources of inherent variation in the speech process. *Invariance and Variability in Speech processes*, pp. 402-423, Lawrence Erlbaum, 1986.
- [11] M. Clifford. *Microphones*. Blue Ridge Summit, 3rd Edition, 1986.
- [12] Gonzales, R. and Wintz, P. (1987). *Digital Image Processing*. Addison-Wesley.
- [13] Comon, P. (1994). Independent component analysis—a new concept? *Signal Processing*, 36:287–314.
- [14] Hyvärinen, A. (1998). Independent component analysis in the presence of gaussian noise by maximizing joint likelihood. *Neurocomputing*, 22:49–67.
- [15] Hyvärinen, A. (1999). Gaussian moments for noisy independent component analysis. *IEEE Signal Processing Letters*, 6(6):145–147.
- [16] Delfosse, N. and Loubaton, P. (1995). Adaptive blind separation of independent sources: a deflation approach. *Signal Processing*, 45:59–83.

- [17] Huber, P. (1985). Projection pursuit. *The Annals of Statistics*, 13(2):435–475.
- [18] Cover, T. M. and Thomas, J. A. (1991). *Elements of Information Theory*. Wiley.
- [19] Papoulis, A. (1991). *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, 3rd edition.
- [20] Hyvärinen, A. (1999). Survey on independent component analysis. *Neural Computing Surveys*, 2:94–128.
- [21] Jones, M. and Sibson, R. (1987). What is projection pursuit? *J. of the Royal Statistical Society, Ser. A*, 150:1–36.
- [22] Pham, D.-T., Garrat, P., and Jutten, C. (1992). Separation of a mixture of independent sources through a maximum likelihood approach. In *Proc. EUSIPCO*, pages 771–774.
- [23] Bell, A. and Sejnowski, T. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7:1129–1159.
- [24] Nadal, J.-P. and Parga, N. (1994). Non-linear neurons in the low noise limit: a factorial code maximizes information transfer. *Network*, 5:565–581.
- [25] Pearlmutter, B. A. and Parra, L. C. (1997). Maximum likelihood blind source separation: A context-sensitive generalization of ICA. In *Advances in Neural Information Processing Systems*, volume 9, pages 613–619.
- [26] Cardoso, J.-F. and Laheld, B. H. (1996). Equivariant adaptive source separation. *IEEE Trans. on Signal Processing*, 44(12):3017–3030.
- [27] Hyvärinen, A. and Oja, E. (1998). Independent component analysis by general nonlinear Hebbian-like learning rules. *Signal Processing*, 64(3):301–313.
- [28] Lee, T.-W., Girolami, M., and Sejnowski, T. J. (1999). Independent component analysis using an extended infomax algorithm for mixed sub-gaussian and super-gaussian sources. *Neural Computation*, 11(2):417–441.
- [29] Hyvärinen, A., Särelä, J., and Vigário, R. (1999). Spikes and bumps: Artefacts generated by independent component analysis with insufficient sample size. In *Proc. Int. Workshop on Independent Component Analysis and Signal Separation (ICA'99)*, pages 425–429, Aussois, France.

- [30] Hyvärinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, 10(3):626–634.
- [31] Luenberger, D. (1969). Optimization by Vector Space Methods. Wiley.
- [32] Karhunen, J., Oja, E., Wang, L., Vigário, R., and Joutsensalo, J. (1997). A class of neural networks for independent component analysis. *IEEE Trans. on Neural Networks*, 8(3):486–504.
- [33] Amari, S.-I., Cichocki, A., and Yang, H. (1996). A new learning algorithm for blind source separation. In *Advances in Neural Information Processing Systems 8*, pages 757–763. MIT Press.
- [34] Cichocki, A. and Unbehauen, R. (1996). Robust neural networks with on-line learning for blind identification and blind separation of sources. *IEEE Trans. on Circuits and Systems*, 43(11):894–906.
- [35] Hyvärinen, A. (1999b). The fixed-point algorithm and maximum likelihood estimation for independent component analysis. *Neural Processing Letters*, 10(1):1–5.
- [36] Vigário, R., Jousmäki, V., Hämäläinen, M., Hari, R., and Oja, E. (1998). Independent component analysis for identification of artifacts in magnetoencephalographic recordings. In *Advances in Neural Information Processing Systems*, volume 10, pages 229–235. MIT Press.
- [37] Vigário, R., Särelä, J., and Oja, E. (1998). Independent component analysis in wave decomposition of auditory evoked fields. In *Proc. Int. Conf. on Artificial Neural Networks (ICANN'98)*, pages 287–292, Skövde, Sweden.
- [38] Back, A. D. and Weigend, A. S. (1997). A first application of independent component analysis to extracting structure from stock returns. *Int. J. on Neural Systems*, 8(4):473–484.
- [39] Kiviluoto, K. and Oja, E. (1998). Independent component analysis for parallel financial time series. In *Proc. Int. Conf. on Neural Information Processing (ICONIP'98)*, volume 2, pages 895–898, Tokyo, Japan.
- [40] Olshausen, B. A. and Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381:607–609.
- [41] Donoho, D. L., Johnstone, I. M., Kerkycharian, G., and Picard, D. (1995). Wavelet shrinkage: asymptopia? *Journal of the Royal Statistical Society, Ser. B*, 57:301–337.
- [42] Hyvärinen, A. (1999d). Sparse code shrinkage: Denoising of nongaussian data by maximum likelihood estimation. *Neural Computation*, 11(7):1739–1768.

- [43] Mallat, S. G. (1989). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 11:674–693.
- [44] Ristaniemi, T. and Joutsensalo, J. (1999). On the performance of blind source separation in CDMA downlink. In *Proc. Int. Workshop on Independent Component Analysis and Signal Separation (ICA'99)*, pages 437–441, Aussois, France.
- [45] A. Westner. *Object-Based Audio Capture: Separating Acoustically-Mixed Signals*. MSc Thesis, MIT, 1998.
- [46] L. Zhang, S. Amari, A. Cichocki. Natural gradient approach to blind separation of over- and under-complete mixtures. *Proc. International Workshop on ICA and BSS*, pp. 455-460, 1999.
- [47] M. Joho, H. Mathis, R. Lambert. Overdetermined Blind Source Separation: Using more Sensors than source signals in a noisy mixture. *Proc. International Workshop on ICA and BSS*, pp. 81-86, 2000.
- [48] M. S. Pedersen, J. Larsen, U. Kjems, and L. C. Parra, "A survey of convolutive blind source separation methods," in *Springer Handbook of Speech Processing*. New York: Springer, 2007.
- [49] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique using second order statistics," *IEEE Trans. Signal Process.*, vol. 45, no. 2, pp. 434–444, Feb. 1997.
- [50] K. Rahbar and J.-P. Reilly, "A frequency domain method for blind source separation of convolutive audio mixtures," *IEEE Trans. Speech Audio Process.* vol. 13, no. 5, pp. 832–844, May 2005 [Online]. Available: <http://www.ece.mcmaster.ca/~reilly/kamran/id18.htm>
- [51] L. Parra and C. Spence, "Convolutive blind separation of non-stationary sources," *IEEE Trans. Speech Audio Process.* vol. 8, no. 3, pp. 320–327, May 2000, [Online]. Available: http://ida.first.fhg.de/~harmeli/download/download_convbss.html
- [52] D.-T. Pham, C. Servière, and H. Boumaraf, "Blind separation of speech mixtures based on nonstationarity," in *Proc. ISSPA'03*, 2003, vol. 2, pp. 73–76 [Online]. Available: http://www.lis.inpg.fr/pages_perso/bliss/toolboxes/bssaudio-demo.tar.gz, [Online].
- [53] L. De Lathauwer and J. Castaing, "Blind identification of underdetermined mixtures by simultaneous matrix diagonalization," *IEEE Trans. Signal Process.*, vol. 56, no. 3, pp. 1096–1105, May 2008.

- [54] H. Sawada, R. Mukai, S. Araki, and S. Makino, "A robust and precise method for solving the permutation problem of frequency-domain blind source separation," *IEEE Trans. Speech Audio Process.*, vol. 12, no. 5, pp. 530–538, Sep. 2004.
- [55] H. Sawada, S. Araki, and S. Makino, "MLSP 2007 data analysis competition: Frequency-domain blind source separation for convolutive mixtures of speech/audio signals," in *Proc. MLSP'07*, 2007, pp. 45–50.
- [56] D. Nion and N. D. Sidiropoulos, "Adaptive algorithms to track the PARAFAC decomposition of a third-order tensor," *IEEE Trans. Signal Process.*, vol. 57, no. 6, pp. 2299–2310, Jun. 2009.
- [57] J. B. Kruskal, "Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics," *Linear Algebra Appl.*, vol. 18, pp. 95–138, 1977.
- [58] A. Stegeman and N. D. Sidiropoulos, "On Kruskal's uniqueness condition for the CANDECOMP/PARAFAC decomposition," *Linear Algebra Appl.*, vol. 420, pp. 540–552, 2007.
- [59] A. Stegeman, J. ten Berge, and L. De Lathauwer, "Sufficient conditions for uniqueness in CANDECOMP/PARAFAC and INDSCAL with random component matrices," *Psychometrika*, vol. 71, pp. 219–229, 2006.
- [60] R. A. Harshman, "Foundations of the PARAFAC procedure: Model and conditions for an 'explanatory' multi-mode factor analysis," *UCLA Working Papers in Phonetics*, vol. 16, pp. 1–84, 1970.
- [61] N. D. Sidiropoulos, G. B. Giannakis, and R. Bro, "Blind PARAFAC receivers for DS-CDMA systems," *IEEE Trans. Signal Process.*, vol. 48, no. 3, pp. 810–823, Mar. 2000.
- [62] R. Bro, "PARAFAC: Tutorial and applications," *Chemom. Intell. Lab. Syst.*, vol. 38, pp. 149–171, 1997.
- [63] M. Rajih and P. Comon, "Enhanced line search: A novel method to accelerate PARAFAC," in *Proc. Eusipco'05*, 2005.
- [64] D. Nion and L. De Lathauwer, "An enhanced line search scheme for complex-valued tensor decompositions. Application in DS-CDMA," *Signal Process.*, vol. 88, no. 3, pp. 749–755, 2008.
- [65] G. Tomasi and R. Bro, "A comparison of algorithms for fitting the PARAFAC model," *Comput. Statist. Data Anal.*, vol. 50, pp. 1700–1734, 2006.

- [66] L. De Lathauwer, "A link between the canonical decomposition in multilinear algebra and simultaneous matrix diagonalization," *SIAM J. Matrix Anal. Appl.*, vol. 28, no. 3, pp. 642–666, 2006.
- [67] L. De Lathauwer and J. Castaing, "Tensor-based techniques for the blind separation of DS-CDMA signals," *Signal Process., Special Iss. Tensor Signal Process.*, vol. 87, no. 2, pp. 322–336, 2007.
- [68] L. De Lathauwer and J. Vandewalle, "Dimensionality reduction in higher-order signal processing and rank- $(r_1, r_2, \dots, r_n)$ reduction in multilinear algebra," *Linear Algebra Appl., Special Iss. Linear Algebra Signal Image Process.*, vol. 391, pp. 31–55, Nov. 2004.
- [69] A.-J. van der Veen and A. Paulraj, "An analytical constant modulus algorithm," *IEEE Trans. Signal Process.*, vol. 44, pp. 1136–1155, May 1996.
- [70] R. Mukai, H. Sawada, S. Araki, and S. Makino, "Frequency-domain blind source separation of many speech signals using near-field and far-field models," *EURASIP J. Appl. Signal Process.*, vol. 2006, pp. 1–13, 2006.
- [71] K. Matsuoka and S. Nakashima, "Minimal distortion principle for blind source separation," in *Proc. Int. Workshop Ind. Compon. Anal. Blind Signal Separation (ICA'01)*, 2001, pp. 722–727.
- [72] J. Anemüller and B. Kollmeier, "Amplitude modulation decorrelation for convolutive blind source separation," in *Proc. Int. Workshop Ind. Compon. Anal. Blind Signal Separation (ICA'00)*, 2000, pp. 215–220.
- [73] N. Murata, S. Ikeda, and A. Ziehe, "An approach to blind source separation based on temporal structure of speech signals," *Neurocomputing*, vol. 41, pp. 1–24, Oct. 2001.
- [74] N. Mitianoudis and M. Davies, "Audio source separation of convolutive mixtures," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 5, pp. 489–497, Sep. 2003.
- [75] R. Aichner, H. Buchner, S. Araki, and S. Makino, "On-line time-domain blind source separation of nonstationary convolved signals," in *Proc. Int. Workshop Indep. Comp. Anal. Blind Sig. Separation (ICA'03)*, 2003, pp. 987–992.
- [76] L. Parra and C. Spence, "On-line convolutive source separation of nonstationary signals," *J. VLSI Signal Process.*, vol. 26, no. 1–2, Aug. 2000.