ΑΝΑΓΝΩΡΙΣΗ ΜΟΥΣΙΚΟΥ ΕΙΔΟΥΣ: ΜΙΑ ΒΙΟ-ΕΜΠΝΕΥΣΜΕΝΗ ΠΟΛΥΓΡΑΜΜΙΚΗ ΠΡΟΣΕΓΓΙΣΗ

Μεταπτυχιακή Διατριβή

ΙΩΑΝΝΗ Κ. ΠΑΝΑΓΑΚΗ

Πτυχιούχου του Τμήματος Πληροφορικής και Τηλεπικοινωνιών, Ε.Κ.Π.Α.

Επιβλέπων: Κωνσταντίνος Κοτρόπουλος, Αναπληρωτής Καθηγητής

ΑΡΙΣΤΟΤΕΛΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΟΝΙΚΗΣ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ - ΚΑΤΕΥΘΥΝΣΗ ΨΗΦΙΑΚΩΝ ΜΕΣΩΝ

ΑΚΑΔΗΜΑΪΚΟ ΕΤΟΣ 2007-2008

Μέλη τριμελούς επιτροπής: Αν. Καθηγητής Κωνσταντίνος Κοτρόπουλος, Λέκτορας Νικόλαος Λάσκαρης, και Καθηγητής Ιωάννης Πήτας.

Στην Ξένια

Abstract

In this master thesis, automatic musical genre classification is addressed under a multilinear perspective. Inspired by a model of auditory cortical processing, multiscale temporal and spectro-temporal modulation features are extracted. Recently, such temporal and spectro-temporal modulation features have been successfully used in various content-based audio classification tasks, but not yet in musical genre classification. Each recording is represented by a N-order feature tensor generated by the auditory model. Thus, each ensemble of recordings is represented by a $(N+1)$-order data tensor created by stacking the N-order feature tensors associated to the recordings, where $N = 2$ or $N = 3$ according to type o modulation features are used. To handle large data tensors and derive compact feature vectors suitable for classification, two new Non-negative Tensor Factorization Algorithms (NTF) are proposed, namely the Projected Landweber -NTF and Coordinate Wise-NTF based on Least Squares Error (LSE) minimization by employing projected gradient techniques. These algorithms guarantee that the limit point of the optimization is a stationary point. Additionally, three other multilinear subspace analysis techniques are employed as multilinear feature extraction techniques, namely a classical Non-Negative Tensor Factorization, the High-Order Singular Value Decomposition (HOSVD), and the Multilinear Principal Component Analysis (MPCA). Classification is performed by a Support Vector Machine (SVM) and a Nearest Neighbour (NN) classifier. Stratified cross-validation tests on two well known datasets, namely the GTZAN dataset and the ISMIR 2004 GENRE one, demonstrate the advantages of the proposed NTF algorithms among the other multilinear subspace analysis methods. The effectiveness of the proposed approach exceeds the accuracies achieved by the state-of-the-art music genre classification algorithms and is near 83.%.

Πρόλογος

Η εργασία αυτή αποτελεί την μεταπτυχιακή διατριβή του Ιωάννη Παναγάκη. Εκπονήθηκε στο εργαστήριο Τεχνητής Νοημοσύνης και Ανάλυσης Πληροφοριών του Τμήματος Πληροφορικής του Αριστοτελείου Πανεπιστημίου Θεσσαλονίκης κατά το ακαδημαϊκό έτος 2007-2008, υπό την επίβλεψη του Αναπληρωτή Καθηγητή του Τμήματος Πληροφορικής του Α.Π.Θ. κ. Κωνσταντίνο Κοτρόπουλο. Θα ήθελα να ευχαριστήσω θερμά τον επιβλέποντα καθηγητή του μεταπτυχιακού μου Αναπληρωτή Καθηγητή κ. Κωνσταντίνο Κοτρόπουλο για την στενή συνεργασία, τις γνώσεις, τις ιδέες, την ενθάρρυνση και τις ουσιαστικές συμβουλές, που απλόχερα μου προσέφερε, καθ' όλη την διάρκεια των μεταπτυχιακών μου σπουδών στο Τμήμα Πληροφορικής του Αριστοτελείου Πανεπιστημίου Θεσσαλονίκης. Επίσης θα ήθελα να εκφράσω τις ευχαριστίες μου και στα υπόλοιπα μέλη Δ.Ε.Π. της κατεύθυνσης Ψηφιακών Μέσων, τον Καθηγητή κ. Ιωάννη Πήτα, τον Επίκουρο Καθηγητή κ. Νικόλαο Νικολαϊδη και τον Λέκτορα κ. Νικόλαο Λάσκαρη, καθώς και σε όλα τα μέλη του εργαστηρίου που μοιραστήκαμε πολλά τον τελευταίο χρόνο.

Θεσσαλονίκη, Ιούνιος του 2008

Περιεχόμενα

Κεφάλαιο 1

Εισαγωγή

1.1 Στόχοι Διατριβής

Είναι γεγονός, ότι την τελευταία δεκαετία με την εξέλιξη των δικτύων υπολογιστών, των υπηρεσιών διαδικτύου καθώς και των κατανεμημένων αρχιτεκτονικών, όπως τα δίκτυα ομότιμων κόμβων (peer-to-peer networks) το διαδίκτυο καθίσταται ιδανικό μέσο για τη διακίνηση πολυμεσικών αρχείων και ιδιαίτερα αρχείων μουσικής. Ως εκ τούτου, ολοένα και αυξάνεται το πλήθος αρχείων μουσικής που αποθηκεύονται τόσο σε προσωπικούς υπολογιστές, όσο και σε εξυπηρετητές ηλεκτρονικών καταστημάτων και ψηφιακών βιβλιοθηκών. Η οργάνωση και η αποδοτική διαχείριση των αρχείων μουσικής είναι μείζονος σημασίας για την ορθή λειτουργία τέτοιων συστημάτων. Μολονότι έχουν αναπτυχθεί εργαλεία για την χειροκίνητη οργάνωση των αρχείων, ο όγκος τους είναι τόσο μεγάλος που καθιστά το εγχείρημα σχεδόν αδύνατο. Η ανάγκη ανάπτυξης εργαλείων αυτόματης ταξινόμησης, αναζήτησης, και ανάκτησης των αρχείων κρίνεται επιτακτική.

Τέτοιου είδους εργαλεία αυτόματης ταξινόμησης πρέπει να μπορούν να εξαγάγουν χρήσιμες πληροφορίες για τα μουσικά κομμάτια απευθείας από το ψηφιακό σήμα. Πιθανές πληροφορίες περιγραφής του περιεχομένου του μουσικού κομματιού θα μπορούσε να είναι: το μουσικό είδος (music genre), η διάθεση (mood), το μουσικό στυλ (music style), ο εκτελεστής - ερμηνευτής του μουσικού κομματιού [78, 84]. Ο Aucouturier και ο Pachet [3] υποστηρίζουν ότι το μουσικό είδος είναι πιθανώς η δημοφιλέστερη περιγραφή του περιεχομένου ενός μουσικού τραγουδιού.

Η ταξινόμηση των καταγραφών μουσικής σε διακριτά είδη είναι ένα δημοφιλές και ανοιχτό ερευνητικό πεδίο στην επονομαζόμενη ερευνητική κοινότητα *Ανάκτησης Μουσικής Πληροφορίας* (Music Information Retrieval-MIR). Παρόλα αυτά, το πρόβλημα της αναγνώρισης μουσικού είδους δεν είναι τετριμμένο: είναι χαρακτηριστικό ότι δεν έχει προταθεί καμία γενική ταξινόμηση για μουσικά είδη και ακόμα και η ανθρώπινη ακρίβεια ταξινόμησης φτάνει το 72% για μη ειδικούς

στο πεδίο.

Στη βιβλιογραφία έχουν προταθεί αρκετά συστήματα αναγνώρισης μουσικού είδους, τα οποία εξάγουν χαρακτηριστικά διανύσματα χροιάς, χρονικά-ρυθμικά χαρακτηριστικά, χαρακτηριστικά που στηρίζονται στη θεμελιώδη (συχνότητα μουσικού τόνου - pitch) καθώς και μελωδικά-αρμονικά χαρακτηριστικά και χρησιμοποιούν κάποιο αλγόριθμο μηχανικής μάθησης για να εκτελέσει την ταξινόμηση των αρχείων κατόπιν εκπαίδευσης. Μία άλλη λογική είναι η ομαδοποίηση των αρχείων με βάση τα χαρακτηριστικά περιγραφής τους, αφού ούτως ή άλλως δεν υπάρχει κάποια κοινώς αποδεκτή ταξινόμηση των μουσικών ειδών.

Πρόσφατα, στους κόλπους της ερευνητικής κοινότητας που ασχολείται με την Ανάκτηση Μουσικής Πληροφορίας, εκφράστηκε ο προβληματισμός ότι το μουσικό είδος είναι διφορούμενος και ασυμβίβαστος τρόπος οργάνωσης και αναζήτησης μουσικής, και ως εκ τούτου οι ανάγκες των χρηστών θα καλύπτονταν καλύτερα με τη χρήση περιγραφών βασισμένων στη μουσική ομοιότητα [86]. Από μια διαφορετική σκοπιά στο [75], επισημαίνεται ότι οι τελικοί χρήστες είναι πιθανότερο να αναζητήσουν μουσική με βάση το μουσικό είδος παρά με βάση την ομοιότητα καλλιτεχνών ή την ομοιότητα μουσικής. Επιπλέον, ο Aucouturier στο [4] διαπιστώνει ότι τα πρόσφατα συστήματα αναγνώρισης μουσικού είδους, απέτυχαν να αυξήσουν σημαντικά την απόδοσης τους, εν συγκρίσει με τα πρόωρα συστήματα και ότι τα ποσοστά ορθής ταξινόμησης μουσικού είδους τα καθιστούν μη αξιόπιστα, ώστε να χρησιμοποιηθούν στην πράξη. Είναι σαφές ότι οι νέες προσεγγίσεις απαιτούνται για να καταστήσουν τα συστήματα αυτόματης αναγνώρισης μουσικού είδους βιώσιμα στην πράξη. Ο McKay στο [86] επιχειρηματολογεί υπέρ της συνέχισης της έρευνας στο πεδίο της αυτόματη αναγνώρισης μουσικού είδους και ενθαρρύνει την κοινότητα MIR να προσεγγίσει διεπιστημονικά το πρόβλημα.

Στην παρούσα διατριβή το πρόβλημα της ταξινόμησης μουσικού είδους προσεγγίζεται υπό μια νέα οπτική. Πρώτος στόχος αυτής της νέας προσέγγισης είναι η συστηματοποίηση, και η θεωρητική θεμελίωση των χαρακτηριστικών που εξάγονται και χρησιμοποιούνται για την αναγνώριση μουσικού είδους. Δεύτερος στόχος αποτελεί η εξαγωγή ορθής αναπαράστασης χαμηλής τάξης (low rank) των χαρακτηριστικών αυτών χρησιμοποιώντας κατάλληλες πολυγραμμικές τεχνικές. Η αποσαφήνιση των στόχων ακολουθεί.

Τα αποτελέσματα ψυχοφυσιολογικών και ψυχοακουστικών ερευνών [109, 133], των τελευταίων σαράντα ετών, καταδεικνύουν τη σημασία των χρονικών (temporal) και χρονοφασματικών (spectro-temporal) διαμορφώσεων του ηχητικού σήματος για την αντίληψη και ερμηνεία του ήχου από τον άνθρωπο. Παρακινούμενοι από την παραπάνω διατύπωση, χρησιμοποιούμε υπολογιστικά μοντέλα [83, 130] του ανθρώπινου ακουστικού συστήματος σε συνδυασμό με προηγμένες τεχνικές επεξεργασίας σήματος έτσι ώστε να εξάγουμε βιο-εμπνευσμένες αναπαραστάσεις που απεικονίζουν ένα δεδομένο μονοδιάστατο ακουστικό σήμα σε ένα χώρο πολλών διαστάσεων,

που περιγράφει το λανθάνον περιεχόμενο των φασματικών και χρονικών διαμορφώσεών του. Για κάθε αρχείο μουσικής εξάγονται αναπαραστάσεις του μουσικού σήματος που περιγράφουν το περιεχόμενο μόνο των χρονικών διαμορφώσεων, και αναπαραστάσεις που περιγράφουν το από κοινού περιεχόμενο χρονικών και φασματικών διαμορφώσεων. Οι προαναφερθήσες αναπαραστάσεις, στην παρούσα διατριβή, θα αναφερονται γενικά ως *αναπαράστασεις φλοιού*, μια και μιμούνται τον τρόπο αναπαράστασης του ήχου στον ακουστικό φλοιό του εγκεφάλου. Η μαθηματική αναπαράσταση των χρονικών διαμορφώσεων μπορεί να γίνει χρησιμοποιώντας πίνακες ή τανυστές δεύτερης τάξης, ενώ η μαθηματική αναπαράσταση των χρονικών-συχνοτικών διαμορφώσεων χρησιμοποιώντας ένα τρισδιάστατο πίνακα ή τανυστή τρίτης τάξης. Συνεπώς, μπορούμε να θεωρήσουμε ότι αυτές οι πολυδιάστατες αναπαραστάσεις φλοιού ορίζονται σε έναν τανυστικό χώρο μεγάλων διαστάσεων. Η αναπαράσταση ενός συνόλου από αρχεία μουσικής γίνεται από *τανυστές δεδομένων* τρίτης και τέταρτης τάξης για κάθε αναπαράσταση φλοιού αντίστοιχα.

Η απευθείας χρήση των παραπάνω αναπαραστάσεων ως είσοδος σε κάποιον αλγόριθμο μηχανικής μάθησης, λόγω των μεγάλων διαστάσεών τους, πάσχει από την αποκαλούμενη *κατάρα των μεγάλων διαστάσεων* (*curse of dimensionality*) [38, 80]. Επιπλέον, αυτές οι αναπαραστάσεις φλοιού είναι ιδιαίτερα πλεοναστικές (redundant). Επομένως, είναι λογικό να θεωρήσουμε ότι οι τανυστές δεδομένων μπορούν να περιγραφούν σ΄ ένα χώρο λιγότερων διαστάσεων χωρίς ουσιαστική απώλεια πληροφορίας. Τεχνικές *μείωσης των διαστάσεων* (dimensionality reduction) που έχουν ως στόχο να μετασχηματίσούν έναν τέτοιο χώρο πολλών διαστάσεων σε ένα χώρο λιγότερων διαστάσεων, διατηρώντας το μεγαλύτερο μέρος της πληροφορίας σχετικά με τη λανθάνουσα δομή της πραγματικής αναπαράστασης φλοιού είναι επιθυμητές.

Σύμφωνα με υπολογιστικές θεωρίες, που ερμηνεύουν τη διαδικασία της αντίληψης του ήχου από τον εγκέφαλο, η αντίληψη του όλου βασίζεται στην αντίληψη των μερών του. Δηλαδή, οι ακουστικές πληροφορίες στον εγκέφαλο αναπαρίστανται σαν γραμμικός συνδυασμός βασικών στοιχείων πληροφορίας (συναρτήσεις βάσης), και προσθετικοί συνδυασμοί πολλών συναρτήσεων βάσης αναπαριστούν την πληροφορία [28]. Η διαπίστωση αυτή μας παρακινεί να χρησιμοποιήσουμε την πολυγραμμική τεχνική ανάλυσης υποχώρων της *Παραγοντοποίησης Μη-αρνητικών Τανυστών* (Non-Negative Tensor Factorization) ως μέθοδο για την μείωση των διαστάσεων και την εξαγωγή καταλλήλων μη ολιστικών χαρακτηριστικών από τις αναπαραστάσεις φλοιού.

Οι μέθοδοι παραγοντοποίησης μη αρνητικών τανυστών που έχουν προταθεί στην βιβλιογραφία συχνά περιορίζονται σε παραγοντοποιήση τανυστών μέχρι τρίτης τάξης και στερούνται αποδείξεων που εγγυώνται τη σύγκλιση τους σε στάσιμο σημείο. Στην παρούσα μεταπτυχιακή διατριβή προτείνονται δυο νέοι αλγόριθμοι παραγοντοποίησης μη αρνητικών τανυστών χρησιμοποιώντας τεχνικές προβεβλημένων διανυσμάτων κλίσης (projected gradients). Οι δύο νέοι αλγόριθμοι έχουν θεωρητικά θεμελιωμένες ιδιότητες σύγκλισης και μπορούν να εφαρμοστούν

σε τανυστές N τάξης συμπεριλαμβανομένων και των τανυστών για $N = 2$, αντιμετωπίζοντας έτσι την παραγοντοποίηση μη αρνητικών πινάκων (Non Negative Matrix Factorization - NMF) ως υποπερίπτωση της παραγοντοποιήσης μη αρνητικών τανυστών.

Για λόγους πληρότητας και σύγκρισης, εκτός από τις προτεινόμενες μεθόδους NTF χρησιμοποιούμε και με μια σειρά άλλων πολυγραμμικών μεθόδων ανάλυσης υποχώρων, που έχουν πρόσφατα προταθεί στην βιβλιογραφία, για την εξαγωγή χαρακτηριστικών από τις ηχιτικές αναπαραστάσεις φλοιού. Πιο συγκεκριμένα χρησιμοποιούμε, την Πολυγραμμική Ανάλυση Πρωτευουσών Συνιστωσών (Multilinear Principal Component Analysis - MPCA) [80], την Υψηλής Τάξης Αποσύνθεση Ιδιαζουσών Τιμών (High Order Singular Value Decomposition - HOSVD) [72] [73] καθώς και τη μέθοδο NTF που προτάθηκε από τον Μπενέτο στο [8, 9]. Για την αναγνώριση μουσικού είδους χρησιμοποιούμε ταξινομητές πλησιέστερου γείτονα καθώς και Μηχανές Εδραίων Διανυσμάτων (Support Vector Machines-SVMs).

Εκτεταμένα πειράματα σε δύο γνωστά σύνολα δεδομένων, το σύνολο δεδομένων GTZAN και το σύνολο δεδομένων ISMIR2004Genre, καταδεικνύουν τα πλεονεκτήματα των προτεινόμενων αλγορίθμων NTF μεταξύ των άλλων πολυγραμμικών μεθόδων ανάλυσης υποχώρων. Η ακρίβεια ορθής αναγνώρισης μουσικού είδους της προτεινόμενης προσέγγισης φτάνει και ξεπερνά το 83% στην καλύτερη περίπτωση, ξεπερνώντας οριακά αυτή των καλύτερων αλγορίθμων αναγνώρισης μουσικού είδους που έχουν προταθεί στην βιβλιογραφία.

Αξίζει να σημειωθεί ότι οι προτεινόμενοι αλγόριθμοι είναι γενικοί και μπορούν να χρησιμοποιηθούν και σε άλλα επιστημονικά πεδία, όπως η επεξεργασία εικόνας, βίντεο και αλλού.

1.2 Διάρθρωση Διατριβής

Ακολουθεί η οργάνωση της διπλωματικής σε κεφάλαια:

- Στο Κεφάλαιο 2 ο αναγνώστης, αρχικά, εισάγεται στην ανατομία και φυσιολογία του ανθρώπινου συστήματος ακοής. Στη συνέχεια παρουσιάζονται τα βασικά ψυχοακουστικά και ψυχοφυσιολογικά ευρήματα που καταδεικνύουν τη σημασία των χρονικών και συχνοτικών διαμορφώσεων για την αντίληψη του ήχου από τον άνθρωπο. Το υπολογιστικό μοντέλο του συστήματος ακοής καθώς και αλγόριθμοι για την εξαγωγή των αναπαραστάσεων φλοιού του ήχου παρουσιάζονται στα εδάφια του κεφαλαίου.

- Στο Κεφάλαιο 3 δίνεται μία εισαγωγή στις βασικές έννοιες της γραμμικής και πολυγραμμικής άλγεβρας. Έμφαση δίνεται στον ορισμό των τανυστών και στις ιδιότητές τους.

- Η συνοπτική διερεύνηση του πεδίου της αυτόματης αναγνώρισης μουσικού είδους αποτελεί το περιεχόμενο του Κεφάλαιου 4. Αναφέρονται τα βασικά σύνολα δεδομένων που

χρησιμοποιούνται σε πειράματα, τα χαρακτηριστικά περιγραφής των δεδομένων μουσικού είδους, και οι συνήθεις αλγόριθμοι ταξινόμησης.

- Στο Κεφάλαιο 5 παρουσιάζονται βασικές πολυγραμμικές τεχνικές ανάλυσης υποχώρων οι οποίες στην παρούσα διατριβή θα χρησιμοποιηθούν για την εξαγωγή χαρακτηριστικών από τις αναπαραστάσεις φλοιού του μουσικού είδους. Πιο συγκεκριμένα παρουσιάζονται οι πολυγραμμικές τεχνικές ανάλυσης υποχώρων PARAFAC, MPCA και HOSVD. Στην ενότητα 5.3.4 παρουσίαζεται η πλειονότητα των NTF αλγορίθμων που έχουν προταθεί στη βιβλιογραφία. Οι προτεινόμενοι αλγόριθμοι παραγοντοποιήσης μη αρνητικών τανυστών δίνονται στην Ενότητα 5.3.5. Μελετάται η σύγκλιση τους και προτείνονται συνθήκες τερματισμού τους.

- Τέλος, στο Κεφάλαιο 6 περιγράφεται η πειραματική διαδικασία και τα αποτελέσματα. Παρουσιάζεται η διαδικασία κατασκευής του τανυστή δεδομένων εκπαίδευσης από τις αναπαραστάσεις φλοιού καθώς και η διαδικασία εξαγωγής χαρακτηριστικών από τις αναπαραστάσεις φλοιού με την χρήση πολυγραμμικών τεχνικών ανάλυσης υποχώρων. Τέλος, παρουσιάζονται τα πειραματικά αποτελέσματα μαζί με σχόλια και συμπεράσματα. Οι μελλοντικές ερευνητικές κατευθύνσεις παρουσιάζονται στο τέλος του κεφαλαίου.

Κεφάλαιο 2

Αναπαράσταση Μουσικού Σήματος

2.1 Ανατομία και Φυσιολογία της Ακοής

Το ανθρώπινο σύστημα ακοής διαδραματίζει σπουδαίο ρόλο στον τρόπο που αντιλαμβανόμαστε τον ήχο. Λειτουργώντας ως σύστημα μετατροπής, μετατρέπει την ηχητική γλώσσα του έξω κόσμου, δηλαδή την ακουστική πίεση, σε ηλεκτρικές ώσεις, δηλαδή τη γλώσσα του εγκεφάλου. Το ανθρώπινο σύστημα ακοής απαρτίζεται από τρία τμήματα: *το περιφερειακό* (peripheral), *το ενδιάμεσο* (intermediate) και *το κεντρικό* (central) [94, 122]. Στην παρούσα διατριβή θα θεωρήσουμε μια απλουστευμένη, από ανατομικής και φυσιολογικής άποψης, εκδοχή του συστήματος ακοής, στην οποία το ενδιάμεσο τμήμα παραλείπεται.

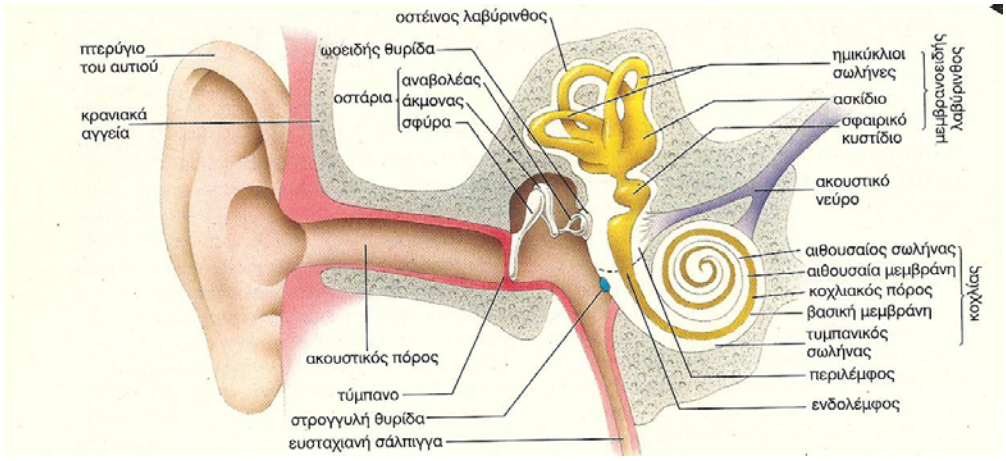
2.1.1 Περιφερειακό Σύστημα Ακοής

Το περιφερειακό τμήμα του ακουστικού συστήματος διαιρείται σε τρία τμήματα: το εξωτερικό, το μέσο και το εσωτερικό αυτί (Σχήμα 2.1).

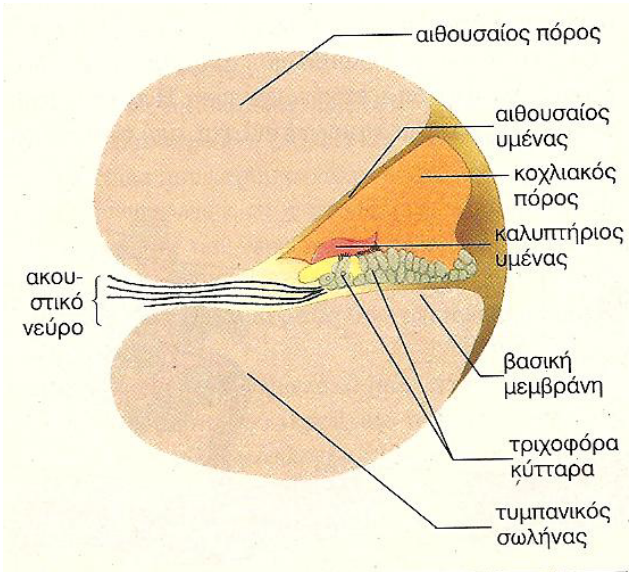
Το εξωτερικό αυτί περιλαμβάνει το πτερύγιο και τον ακουστικό πόρο, δια μέσου του οποίου ο ήχος αφενός συγκεντρώνεται (πτερύγιο) και αφετέρου διαδίδεται (ακουστικός πόρος) στο μέσο αυτί, ενώ παράλληλα φιλτράρεται με ανωδιαβατό φίλτρο στο πτερύγιο και ζωνοδιαβατό φίλτρο στον ακουστικό πόρο.

Το μέσο αυτί, περιλαμβάνει την τυμπανική μεμβράνη καθώς και τα ακουστικά οστάρια: τη σφύρα, τον άκμονα και τον αναβολέα. Τα ακουστικά οστάρια μεταδίδουν τις ταλαντώσεις της τυμπανικής μεμβράνης στο εσωτερικό αυτί. Το εξωτερικό αυτί συν το εσωτερικό μπορούν να μοντελοποιηθούν ως φίλτρο με ευρεία κορυφή 30dB γύρω στα 1000Hz.

Το εσωτερικό αυτί αποτελείται από τον *κοχλία* (Cochlea)(Σχήμα 2.2) και άλλα όργανα, τα οποία ενημερώνουν τον εγκέφαλο για τη θέση του σώματος στο χώρο. Ο κοχλίας είναι ένας σπειροειδής σωλήνας με ολοένα μικρότερη διατομή καθώς τυλίγεται εσωτερικά. Η διατομή μιας



Σχήμα 2.1: Ανατομία του περφεριακού συστήματος ακοής, (από το [135]).



Σχήμα 2.2: Διατομή του κοχλία (από το [135]).

σπείρας του κοχλία φανερώνει την παρουσία ενός διαφράγματος που καλείται *βασική μεμβράνη* (basilar membrane), το οποίο χωρίζει το εσωτερικό του κοχλία σε δύο περίπου ίσα μέρη από τα οποία το κάτω μέρος αποτελεί την τυμπανική κλίμακα ενώ το επάνω υποδιαιρείται στην αιθουσιαία κλίμακα και στον κοχλιακό πόρο. Οι κλίμακες είναι γεμάτες υγρό, την περιλέμφο, ενώ ο κοχλιακός πόρος περιέχει άλλο υγρό την ενδολέμφο. Τα υγρά χρησιμεύουν για τη μετάδοση των ηχητικών κυμάτων προς τη βασική μεμβράνη. Ο ήχος εισέρχεται στην αιθουσιαία κλίμακα μέσω της ωοειδούς θυρίδας στην οποία εφάπτεται ο αναβολέας.

Η βασική μεμβράνη είναι υπεύθυνη για την μετατροπή των μηχανικών ταλαντώσεων, οι οποίες δημιουργούνται από τη δόνηση της ωοειδούς θυρίδας μέσω των ακουστικών οσταρίων, σε ηλεκτρικούς παλμούς. Πιο συγκεκριμένα επάνω στην βασική μεμβράνη υπάρχει το όργανο του

Corti το οποίο περιέχει μια σειρά τριχοειδών κυττάρων (Inner Hair Cells -IHC), οργανωμένων τονοτοπικά, που συνδέονται με τις ίνες του ακουστικού νεύρου και δημιουργούν ηλεκτρικούς παλμούς με συχνότητες έως 20kHz.

Το μήκος της βασικής μεμβράνης είναι περίπου 35mm, ενώ το πλάτος της αυξάνει καθώς μεταβαίνουμε από την βάση προς την κορυφή (appex) του κοχλία, δηλαδή από την ωοειδή θυρίδα στο κέντρο της περιέλιξης του κοχλία και η σκληρότητα της ελαττώνεται. Λόγω αυτών των μηχανικών ιδιοτήτων κατά το μήκος της, η βασική μεμβράνη λειτουργεί σαν φασματικός αναλυτής συντονίζοντας κάθε σημείο της σε ταλάντωση διαφορετικής συχνότητας. Έτσι, κοντά στην ωοειδή θυρίδα συντονίζονται οι υψηλές συχνότητες ενώ κοντά στην κορυφή συντονίζονται οι χαμηλές. Με άλλα λόγια, ένας ήχος υψηλής συχνότητας καθώς διατρέχει τη βασική μεμβράνη πολύ γρήγορα εξασθενεί επειδή συναντά λιγότερο σκληρό μέσο καθώς πλησιάζει τον άπηκα. Τόσο οι χαμηλές όσο και οι υψηλές συχνότητες ενός σύνθετου ήχου φτάνουν στη βασική μεμβράνη ταυτόχρονα δια μέσου των υγρών των δύο κλιμάκων. Έτσι οι χαμηλές συχνότητες δεν μπορούν να διαδοθούν ή να συντονίσουν το σκληρό τμήμα της βασικής μεμβράνης.

2.1.2 Κέντρικό Σύστημα Ακοής

Το κεντρικό σύστημα ακοής περιλαμβάνει τον *πρωτοταγή ακουστικό φλοιό* (primary auditory cortex). Σ' αυτό το τμήμα του ανθρώπινου συστήματος ακοής συντελούνται ανώτερου επιπέδου λειτουργίες που σχετίζονται με διαδικασίες, όπως η αντίληψη (perception), η κατανόηση, η μνήμη και ο προσδιορισμός της θέσης της ηχητικής πηγής στο χώρο [94].

Ο ακουστικός φλοιός αποτελεί την πρωτοταγή μονάδα επεξεργασίας των ακουστικών πληροφοριών από τον εγκεφαλικό φλοιό. Οι ακουστικές πληροφορίες εισέρχονται στον ακουστικό φλοιό κωδικοποιημένες ως νευρικοί παλμοί, εκεί επανακωδικοποιούνται από τα νευρικά κύτταρα του φλοιού σε μια αναπαράσταση που είναι γνωστή ως *φλοιώδης αναπαράσταση* (cortical representation).

Τα νευρικά κύτταρα στον κοχλία είναι οργανωμένα τονοτοπικά, κατ' αναλογία με την τονοτοπική οργάνωση των νευρικών κυττάρων της βασικής μεμβράνης στον κοχλία. Ειδικότερα, τα κύτταρα του ακουστικού φλοιού οργανώνονται τονοτοπικά με τρόπο τέτοιο έτσι ώστε ο άξονας της συχνότητας να σχηματίζει ορθή γωνία με τις λωρίδες ISO-συχνοτήτων που καλύπτουν την επιφάνεια του ακουστικού φλοιού, και διαμορφώνουν μια πλήρη αναπαράσταση των αντιλαμβανόμενων συχνοτήτων [87] επί της νευρικής δομής του ακουστικού φλοιού. Εκτός από την τονοτοπική οργάνωση, τα νευρικά κύτταρα οργανώνονται σε στήλες ανάλογα με την κρουστική τους απόκριση. Παραδείγματος χάριν, υπάρχουν ενδείξεις για την ύπαρξη συστηματικής οργάνωσης των νευρικών κυττάρων ανάλογα με το σχήμα και τη συμμετρία της απόκρισης συχνοτήτων τους. Κύτταρα με απόκριση στενού εύρους ζώνης βρίσκονται γύρω από το κεντρικό

μέσα μέρος του ακουστικού φλοιού, ενώ αντίθετα τα κύτταρα με απόκριση ευρύτερου εύρους ζώνης βρίσκονται πιο κοντά στις άκρες του. Επιπλέον, ο Shamma στο [115] αναφέρει ότι τα κύτταρα του πρωτοταγούς ακουστικού φλοιού οργανώνονται τοπογραφικά με βάση την συμμετρία των κρουστικών τους αποκρίσεων. Τα νευρικά κύτταρα των οποίων η απόκριση συχνότητας τους είναι συμμετρική ως προς την κεντρική συχνότητα βρίσκονται στο κέντρο του ακουστικού φλοιού. Προς τις άκρες, η απόκριση συχνότητας των κυττάρων είναι σταδιακά λιγότερο συμμετρική. Επιπλέον, φαίνεται να υπάρχει μια κιονοειδής (σε στήλες) οργάνωση των νευρικών κυττάρων σύμφωνα με τη στερεοακουστική ευαισθησία (binaural sensivity) τους. Τα κύτταρα που διεγείρονται από τα ερεθίσματα και από τα δυο αυτιά (κύτταρα EE) και τα κύτταρα που διεγείρονται από τα ερεθίσματα από ένα αυτί, αλλά όχι από τα ερεθίσματα που προέρχονται από το άλλο αυτί (EI και IE κύτταρα) βρίσκονται σε χωριστές στήλες [94].

Η μορφολογική και τοπογραφική οργάνωση των κυττάρων υπονοεί ότι ο ακουστικός φλοιός, μέχρι ενός σημείου, διατηρεί έναν περιεκτικό χάρτη στη νευρική δομή του που περιγράφει ποσότητες όπως η συχνότητα, το εύρος ζώνης, η συμμετρία του φάσματος, και η θέση στο χώρο του ήχου. Η παραπάνω διατύπωση υπονοεί ότι η ανάλυση των ακουστικών σημάτων πραγματοποιείται σε όλο το ακουστικό σύστημα από τον κοχλία αυτιού, και τα ενδιάμεσα στάδια μέχρι τον ακουστικό φλοιό, κωδικοποιώντας σταδιακά και επανακωδικοποιώντας τις ηχητικές πληροφορίες κατά τέτοιο τρόπο, ώστε να καταστούν δυνατές ανώτερες εγκεφαλικές λειτουργίες, όπως αυτή της αντίληψης.

Επίπλέον, αξίζει να σημειωθεί οτι οι νευροφυσιολογικές έρευνες της ακοής οδηγούν στο συμπέρασμα ότι η αντίληψη και αποκωδικοποίηση των ηχητικών μηνυμάτων είναι πολύ διαφορετική ανάλογα με το αν στη λειτουργία αυτή συμμετέχει περισσότερο το δεξί ή το αριστερό αυτί. Το δεξί αυτί συνδέεται με το αριστερό ημισφαίριο του εγκεφάλου, εκεί όπου βρίσκεται ένα ειδικό νευρωνικό κέντρο (κέντρο Broca) στο οποίο γίνεται η οργάνωση του λόγου και της γλώσσας. Αυτού του είδους η σύνδεση είναι άμεση και ακριβής. Αντίστοιχα το αριστερό αυτί συνδέεται με το δεξί ημισφαίριο, αλλά όμως εκεί δεν γίνεται η επεξεργασία της γλώσσας, όπως στο αριστερό. Έτσι, λοιπόν, θα πρέπει το ηχητικό ερέθισμα να περάσει μέσω του μεσολοβίου (που συνδέει τα ημισφαίρια του εγκεφάλου) στο δεξί ημισφαίριο. Αυτή δεν είναι μόνο μια αργή σύνδεση, αλλά και ιδιαίτερα αναξιόπιστη.

Για αυτό το λόγο ο Tomatis ονόμασε το δεξί αυτί «αυτί οδηγό», επειδή μπορεί : 1.) να αποκωδικοποιεί ακριβέστερα τα ηχητικά μηνύματα, 2.) να ελέγχει αποτελεσματικότερα την παραγωγή του λόγου, 3.) να εστιάζει καλύτερα την προσοχή του ατόμου. Εάν λοιπόν το δεξί αυτί είναι ηγετικό και κυρίαρχο στη σύλληψη και αποκωδικοποίηση των ηχητικών ερεθισμάτων, υπάρχει δεξιά ακουστική πλευρίωση.

Εάν αντίθετα κυριαρχεί το αριστερό αυτί τότε έχουμε αριστερή ακουστική πλευρίωση, με όλα

τα αρνητικά αποτελέσματα που αυτό συνεπάγεται. Η αριστερή ακουστική πλευρίωση είναι μια προστασία του ατόμου, πρόκειται ουσιαστικά για μια συγκινησιακή άμυνα. Αποφεύγοντας κανείς να χρησιμοποιήσει το συντομότερο δρόμο (δεξί αυτί - αριστερό εγκεφαλικό ημισφαίριο), αλλά επιλέγοντας το μακρύτερο (αριστερό αυτί - δεξί εγκεφαλικό ημισφαίριο- μεσολόβιο-αριστερό εγκεφαλικό ημισφαίριο) δημιουργεί μια χρονική καθυστέρηση και ουσιαστικά μια απόσταση ανάμεσα στον έξω κόσμο και στον εαυτό του, χωρίς βέβαια να πάψει εντελώς να ακούει (αποστασιοποίηση από την πηγή που του δημιουργεί πρόβλημα). Ο μακρύτερος νευρολογικά δρόμος έχει σαν αποτέλεσμα την καθυστέρηση της κατανόησης των πληροφοριών με πιθανές συνέπειες, την απώλεια ενέργειας του ατόμου, την κόπωση, την πτώση της μνήμης, την ελάττωση της συγκέντρωσης, τη διάσπαση προσοχής, και την ανορθογραφία.

2.2 Αντίληψη Χρονικών και Συχνοτικών Διαμορφώσεων

Τα τελευταία σαράντα χρόνια τα αποτελέσματα ψυχοφυσιολογικών και ψυχοακουστικών ερευνών καταδεικνύουν ότι οι πολύ χαμηλών συχνοτήτων *χρονικές και συχνοτικές διαμορφώσεις* (spectro-temporal modulations) του ήχου είναι οι κυριοι φορείς πληροφορίας τόσο για το ηχόχρωμα της μουσικής όσο και για το σήμα ομιλίας. Είναι δε, δυνατό να εξαχθούν από αυτές τις χρονικές και φασματικές διαμορφώσεις ποσοτικά χαρακτηριστικά όπως για παράδειγμα οι φωνοσυντονισμοί (formants) του σήματος ομιλίας [124].

Ο Dudley στο κλασικό πλέον άρθρο του [39], στα 1939 πρώτος παρατήρησε ότι τα σήματα ομιλίας και μουσικής είναι διαδικασίες στενού εύρους ζώνης οι οποίες διαμορφώνουν ένα φέρον μεγαλύτερου εύρους ζώνης. Στα 1971, ο Moler [93] πρώτος αναγνώρισε ότι το σύστημα ακοής των θηλαστικών έχει οξεία ευαισθησία στην αντίληψη διαμόρφωσης πλάτους (amplitude modulation) ή χρονικής διαμόρφωσης (temporal modulation) ακουστικών σημάτων στενού εύρους ζώνης. Ο Suga [123] απέδειξε πως κάθε σημείο της βασικής μεμβράνης αποκρίνεται σε διαφορετικές συχνότητες χρονικών διαμορφώσεων με μια λογική ανάλογη με αυτή που λαμβάνει χώρα κατά την απόκρισή της σε τόνους. Με άλλα λόγια η απόκριση των κοχλιακών κυττάρων στις διάφορες συχνότητες χρονικής διαμόρφωσης έχει τονοτοπικά χαρακτηριστικά. Οι Schreiner και Urbas [112] διαπίστωσαν ότι αντίστοιχα τονοτοπικά χαρακτηριστικά έχουν και τα νευρικά κύτταρα του πρωτοταγούς ακουστικού φλοιού. Πιο πρόσφατα, ο Kowalski και άλλοι [35, 68, 116] απέδειξαν ότι τα νευρικά κύτταρα στον πρωτοταγή ακουστικό φλοιό του εγκεφάλου - το υψηλότερο επίπεδο του πρώιμου ακουστικού συστήματος - είναι πιο ευαίσθητα σε ήχους που συνδυάζουν φασματικές και χρονικές διαμορφώσεις. Η κρουστική απόκριση των κυττάρων του

φλοιού υπολογίστηκε με κατάλληλα πειράματα, τα οποία κατέδειξαν ότι το κεντρικό σύστημα ακοής υλοποιεί από κοινού φασματική-χρονική ανάλυση πολλαπλών κλιμάκων (multiscale) η οποία ουσιαστικά επανακωδικοποιεί (μετασχηματίζει) το ακουστικό φασματογράφημα του ήχου σε επίπεδο χρονικών και φασματικών διαμορφώσεων. Ο Schulze [113] υποστηρίζει ότι η κωδικοποίηση του μουσικού τόνου (pitch) και του ρυθμού ενδεχομένως να εξηγείται χωριστά από συνελικτικά και πολλαπλασιαστικά μοντέλα. Επιπλέον ο Langner, παρατήρησε σε μαγνητοεγκεφαλογραφήματα ότι το συχνοτικό και ρυθμικό (περιοδικό) περιεχόμενο του ήχου αναπαρίστανται με ορθογώνιους χάρτες στον ακουστικό φλοιό του εγκεφάλου [70, 71, 113].

Ευρήματα από την ψυχοακουστική ισχυροποιούν περαιτέρω την επιχειρηματολογία υπέρ της σημασίας των φασματικών-χρονικών διαμορφώσεων στην αντίληψη του ήχου. Ο Houtgast [58] έδειξε ότι η αντίληψη μιας διαμορφώτριας συχνότητας αποκρύπτει (masking) την αντίληψη γειτονικών διαμορφωτριών συχνοτήτων. Περαιτέρω πειράματα των Bacon και Grantham καταδεικνύουν το γεγονός ύπαρξης καναλιών ανίχνευσης της χρονικής διαμόρφωσης αντιστοίχως προς αυτά της ανίχνευσης την φασματικής συχνότητας τα οποία συχνά αποκαλούνται και κρίσιμες ζώνες (critical bands) [5]. Αντίστοιχα φαινόμενα απόκρυψης παρατηρούνται και σε ήχους με περιεχόμενο από κοινού χρονικών και φασματικών διαμορφώσεων [25].

Συμπερασματικά, θα λέγαμε ότι, σε μια πρώτη φάση, το περιφερειακό σύστημα ακοής εκτελεί φασματική ανάλυση, στη συνέχεια το κεντρικό σύστημα ακοής επανακωδικοποιεί τη φασματική πληροφορία με όρους φασματικών και χρονικών διαμορφώσεων. Εφόσον λειτουργίες, όπως η αντίληψη του ήχου, λαμβάνουν χώρα στο κεντρικό σύστημα ακοής και σε υψηλοτέρα τμήματα του ακουστικού συστήματος η ανάλυση που προηγήθηκε ενισχύει την επιχειρηματολογία υπέρ της σημασίας των φασματικών-χρονικών διαμορφώσεων στην αντίληψη του ήχου.

2.3 Μαθηματικό Μοντέλο Ακουστικού Συστήματος

Η πρόοδος ενός συνόλου εφαρμογών, όπως η αυτόματη αναγνώριση ομιλίας και ομιλητή, η συμπίεση και κωδικοποίηση σημάτων ομιλίας και μουσικής κ.α. εξαρτάται σημαντικά από την κατανόηση της λειτουργίας του ανθρώπινου συστήματος ακοής. Η ιδέα της ενσωμάτωσης γνώσης προερχόμενης από τον τρόπο λειτουργίας του ανθρώπινου συστήματος ακοής στην κατασκευή εύρωστων συστημάτων αναγνώρισης ομιλίας και ήχου εν γένει δεν είναι νέα [54]. Η ύπαρξη εύρωστων και αξιόπιστων μαθηματικών-υπολογιστικών μοντέλων του ανθρωπίνου συστήματος ακοής για την αναπαράσταση και την εξαγωγή χαρακτηριστικών από τα ηχητικά σήματα είναι επιτακτική. Στη βιβλιογραφία έχουν προταθεί αρκετά τέτοια υπολογιστικά μοντέλα τα οποία προσομοιώνουν υποσυστήματα του συστήματος ακοής, όπως τον κοχλία [83, 120] και το κεντρικό ακουστικό σύστημα [130].

Στην παρούσα διατριβή θα χρησιμοποιήσουμε το υπολογιστικό του κοχλία που προτάθηκε από τον Lyon [83, 120] ώστε να προσεγγίσουμε την αναπαράσταση χρονικών διαμορφώσεων του ακουστικού φλοιού και το μοντέλο που προτάθηκε στο [130] για να εξάγουμε τις από κοινού χρονικές και συχνοτικές αναπαραστάσεις φλοιού από τα αρχεία ήχου.

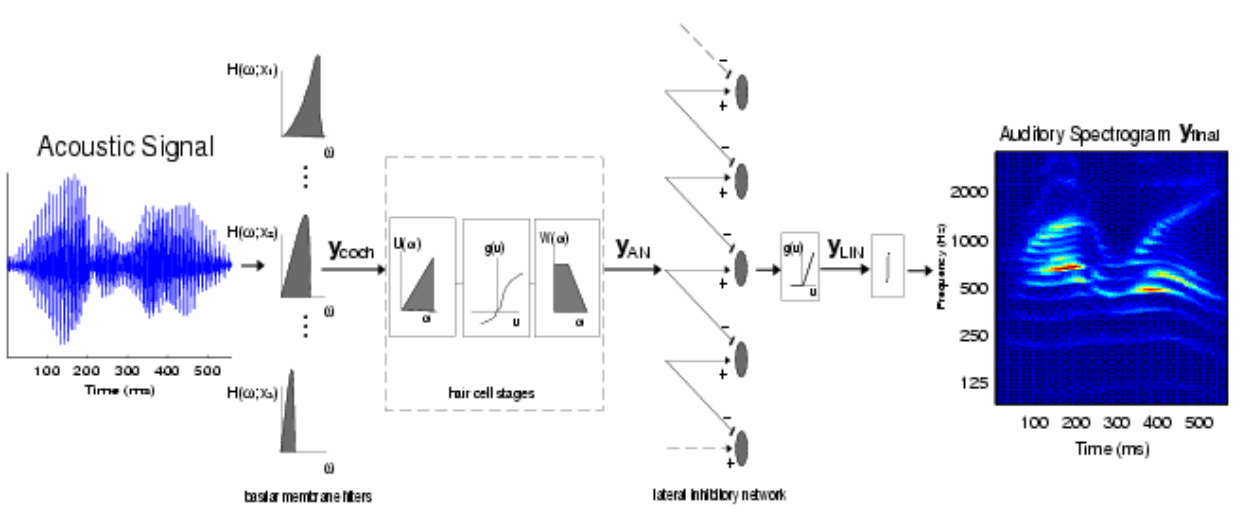
Το ανθρώπινο ακουστικό σύστημα που παρουσιάστηκε στην ενότητα 2.1 μπορεί να μοντελοποιηθεί μαθηματικά ως μια αλληλουχία συστημάτων τα οποία με τη σειρά τους υλοποιούν μια ακολουθία γραμμικών και μη-γραμμικών μετασχηματισμών. Η μαθηματική μοντελοποίηση του ανθρώπινου ακουστικού συστήματος συναντάτε στην βιβλιογραφία ως *Υπολογιστικό Ακουστικό Μοντέλο* (Computational Auditory Model) [88, 103]

Το υπολογιστικό ακουστικό μοντέλο βασίζεται σε νευροφυσιολογικά, βιοφυσικά και ψυχοακουστικά ευρήματα από μελέτες των διάφορων συστατικών μερών του ακουστικού συστήματος και απαρτίζεται από δύο βασικά τμήματα: 1) Το περιφερειακό ακουστικό σύστημα (early stage) το οποίο μοντελοποιεί το μετασχηματισμό του ακουστικού σήματος σε ηλεκτρικές ώσεις. Η αναπαράσταση που προκύπτει στην βιβλιογραφία αναφέρεται ως *ακουστικό φασματογράφημα* (auditory spectrogram). Η λειτουργία του προσομοιώνεται από μια οικογένεια κοχλιακών μοντέλων πού αρχικά προτάθηκε από τον Lyon [83, 120] και τροποποιήθηκε εν μέρη από τους Wang και Shamma για να χρησιμοποιηθεί στη μοντελοποίηση ολοκλήρου του συστήματος ακοής [130]. 2) Το κεντρικό ακουστικό σύστημα (central stage) αναλύει το ακουστικό φασματογράφημα και εκτιμά το περιεχόμενο του σε φασματικές και χρονικές διαμορφώσεις (spectro-temporal modulations), η διαδικασία υλοποιείται από μια τράπεζα φίλτρων τα οποία μιμούνται τη συμπεριφορά των κυττάρων του ανθρώπινου ακουστικού φλοιού (auditory cortex). Η αναπαράσταση που προκύπτει από αυτή την διαδικασία συχνά συναντάται ως αναπαράσταση φλοιού (cortical representation) [26].

2.3.1 Μοντέλο Περιφερειακού Συστήματος Ακοής

Το ακουστικό σήμα $s(t)$ καθώς εισέρχεται στο έσω αυτί δημιουργεί ένα σύνθετο πρότυπο χρονικών και φασματικών διαμορφώσεων κατά μήκος της βασικής μεμβράνης στον κοχλία. Κάθε σημείο του κοχλία συντονίζεται σε ταλάντωση διαφορετικής συχνότητας, συνεπώς η μέγιστη μετατόπιση σε κάθε σημείο του κοχλία προκαλεί την αντίληψη διαφορετικού τόνου, με αποτέλεσμα τη δημιουργία ενός τονοτοπικού άξονα απόκρισης κατά μήκος του κοχλία. Συνεπώς, η βασική μεμβράνη μπορεί να μοντελοποιηθεί ως μια συστοιχία από ζωνοδιαβατά φίλτρα κεντραρισμένα σε ισαπέχουσες συχνότητες επί ένα λογαριθμικό άξονα. Η παραπάνω διαδικασία μπορεί να εκφραστεί μαθηματικά ως *μετασχηματισμός κυματιδίων* (wavelet transform) του ακουστικού σήματος $s(t)$.

Το ακουστικό σήμα $s(t)$ φιλτράρεται από μία συστοιχία φίλτρων στον κοχλία, με κρουστική



Σχήμα 2.3: Ακουστικό φασματογράφημα, (από το [26]).

απόκριση $h_{cochlea}(t, f)$. Η έξοδος τον κοχλιακών φίλτρων $y_{cochlea}(t, f)$ μετατρέπεται μέσω τον τριχοειδών κυττάρων σε ηλεκτρικούς παλμούς. Η παραπάνω διαδικασία μπορεί να περιγραφεί ως μια από κοινού στο χρόνο και στην συχνότητα κατανομή της διέγερσης τριχοειδών κυττάρων, έστω $y_{an}(t, f)$, και δύναται να μοντελοποιηθεί ως μια ακολουθία τριών βημάτων:

1. Αρχικά, το ακουστικό σήμα $s(t)$ φιλτράρεται από ενα ανωδιαβατό φίλτρο με κρουστική απόκριση $h_{cochlea}(t, f)$ (the fluid-cilia coupling), ακολουθούμενο από μια στιγμιαία μη-γραμμική συμπίεση (geted ionic channels) $g_{h_c}(.)$ και ένα κατωδιαβατό φίλτρο (hair cell membrane leakage) $\mu_{h_c(t)}(.)$. Η διαδικασία εκράζεται μαθηματικά ως εξής:

$$y_{cochlea}(t, f) = s(t) * h_{cochlea}(t, f) \quad (2.1)$$

$$y_{an}(t, f) = g_{h_c}(\frac{\partial y_{cochlea}}{\partial t}(t, f)) * \mu_{h_c(t)} \quad (2.2)$$

2. Σε μια δεύτερη φάση ένα πλευρικό ανασταλτικό δίκτυο (Lateral Inhibitory Network - LIN) ανιχνεύει τις ασυνέχειες στις αποκρίσεις κατά μήκος του τονοτοπικού άξονα του ακουστικού νεύρου. Το LIN προσεγγίζεται από την πρώτη παράγωγο ως προς τον τονοτοπικό άξονα ακολουθούμενο από τον ανορθωτή ημίσεος κύματος (half-wave rectifier). Αποτέλεσμα των παραπάνω πράξεων είναι η από κοινού στο χρόνο και στην συχνότητα κατανομή $y_{LIN}(t, f)$ και εκφράζεται μαθηματικά από την σχέση που ακολουθεί.

$$y_{LIN}(t, f) = \max(\frac{\partial y_{an}}{\partial f}(t, f), 0) \quad (2.3)$$

3. Το τελευταίο βήμα αυτής της διαδικασίας είναι η ολοκλήρωση της $y_{LIN}(t, f)$ σε ένα βραχύ χρονικό παράθυρο της τάξης των $8 - 10$ ms γεγονός που μιμείται το χρόνο προσαρμογής

των ακουστικών νευρικών κυττάρων. Το ακουστικό φασματογράφημα $y(t, f)$ προκύπτει ως εξής:

$$y(t, f) = y_{LIN}(t, f) * \mu_{midbrain}(t, \tau) \tag{2.4}$$

όπου με $*$ συμβολίζεται η πράξη της συνέλιξης στο χρόνο.

Η μαθηματική διατύπωση του κοχλιακού μοντέλου που προηγήθηκε αποτελεί γενική διατύπωση της οικογένειας κοχλιακών μοντέλων που προτάθηκε από τον Lyon [83, 120]. Διάφορες τροποποήσεις και παραλαγές αυτού εχουν προταθεί και υλοποιηθεί, π.χ. το Auditory Toolbox[1] και το NLS Toolbox[2] για το περιβάλλον προγραμματισμού Matlab.

2.3.2 Μοντέλο Κεντρικού Συστήματος Ακοής - Αναπαράσταση Φλοιού

Τα επόμενα τμήματα του ακουστικού συστήματος αναλύουν περαιτέρω το ακουστικό φασματογράφημα, με στόχο την εκτίμηση του περιεχομένου του σε φασματικές -χρονικές διαμορφώσεις. Πιο συγκεκριμένα, η παραπάνω διαδικασία εκφράζεται μαθηματικά από τον δυσδιάστατο μετασχηματισμό κυματιδίου, με χρήση μιας δισδιάστατης συνάρτησης Gabor.

Η εκτίμηση των χρονοφασματικών διαμορφώσεων του ακουστικού φασματογραφήματος υπολογίζεται μέσω μιας συστοιχίας φίλτρων επιλογής διαμορφώσεων που είναι κεντραρισμένα σε ισαπέχουσες συχνότητες κατά μήκος του τονοτοπικού άξονα. Κάθε φίλτρο συντονίζεται σε ένα εύρος χρονικών διαμορφώσεων, οι οποίες αναφέρονται ως rates, συμβολίζονται με ω και έχουν ως μονάδα μέτρησης το Hz, καθώς και σε ένα εύρος φασματικών διαμορφώσεων οι οποίες αναφέρονται ως scales συμβολίζονται με Ω και μετρούνται σε Κύκλους ανά Οκτάβα. Μια τυπική κρουστική απόκριση ενός Gabor φίλτρου ή κυματιδίου καλείται Φασματικό - Χρονικό Πεδίο Απόκρισης (Spectro-Temporal Response Field - STRF) και μοντελοποιεί την κρουστική απόκριση ενός δεδομένου κυττάρου του ακουστικού φλοιού ανάλογα με τις παραμέτρους.

Έστω μια συστοιχία κατευθυντικών φίλτρων STRFs αποτελούμενη από επλιεκτικά φίλτρα προς τα κάτω $STRF_-$ και από επιλεκτικά φίλτρα προς τα πάνω $STRF_+$. Κάθε STRF είναι μια πραγματική συνάρτηση, αποτέλεσμα του πραγματικού μέρους, του γινομένου μιας μιγαδικής συνάρτησης χρονικής κρουστικής απόκρισης H_{rate} με μία μιγαδική συνάρτηση φασματικής κρουστικής απόκρισης H_{scale}.

$$STRF_+ = \Re\{H_{rate}(t, \omega, \theta) H_{scale}(f, \Omega, \varphi)\} \tag{2.5}$$

$$STRF_- = \Re\{H^*_{rate}(t, \omega, \theta) H_{scale}(f, \Omega, \varphi)\}. \tag{2.6}$$

[1] http://cobweb.ecn.purdue.edu/ malcolm/interval/1998-010/

[2] http://www.isr.umd.edu/Labs/NSL/Software.htm

όπου με $\Re$ συμβολίζεται το πραγματικό μέρος, με $*$ το μιγαδικό συζυγές , καθώς με θ, φ συμβολίζουμε τις χαρακτηριστικές φάσεις οι οποίες καθορίζουν το βαθμό ασυμμετρίας κατά τον άξονα του χρόνου και της συχνότητας αντιστοιχα.

Για την εξαγωγή των αναλυτικών συναρτήσεων H_{rate} και H_{scale} πρέπει να ορίσουμε τις συναρτήσεις $h_{rate}(.)$, $h_{scale}(.)$ οι οποίες αποτελούν τις χρονικές και φασματικές κρουστικές αποκρίσεις και ορίζονται από μια ημιτονοειδή παρεμβολή ανάμεσα σε μια συμμετρική συνάρτηση $h_r(.)$ (δεύτερη παράγωγος μιας Γκαουσιανής) και $h_s(.)$ (συνάρτηση Γάμμα) και των μη συμμετρικών Hilbert μετασχηματισμών τους ως εξής:

$$h_{rate}(t,\omega,\theta) = h_r(t,\omega)\cos\theta + \tilde{h}_r(t,\omega)\sin\theta \quad (2.7)$$

$$h_{scale}(f,\Omega,\varphi) = h_s(f,\Omega)\cos\varphi + \tilde{h}_s(f,\Omega)\sin\varphi. \quad (2.8)$$

όπου με $\tilde{(.)}$ συμβολίζεται ο μετασχηματισμός Hilbert. Οι κρουστικές αποκρίσεις για διαφορετικές τιμές των παραμέτρων ω και Ω υπολογίζονται ως εξής:

$$h_r(t,\omega) = \omega h_r(\omega t) \quad (2.9)$$

$$h_s(f,\Omega) = \Omega h_s(\Omega f) \quad (2.10)$$

όπου

$$h_r(t) = t^3 e^{-4t}\cos(2\pi t) \quad (2.11)$$

$$h_s(f) = (1-f^2)e^{-\frac{f^2}{2}} \quad (2.12)$$

Συνεπώς οι αναλυτικές συναρτήσεις H_{rate} και H_{scale} προκύπτουν από τις $h_{scale}(.)$ και $h_{rate}(.)$ ως εξής:

$$H_{rate}(t,\omega,\theta) = h_{rate}(t,\omega,\theta) + j\tilde{h}_{rate}(t,\omega,\theta) \quad (2.13)$$

$$H_{scale}(f,\Omega,\varphi) = h_{scale}(f,\Omega,\varphi) + j\tilde{h}_{scale}(f,\Omega,\varphi). \quad (2.14)$$

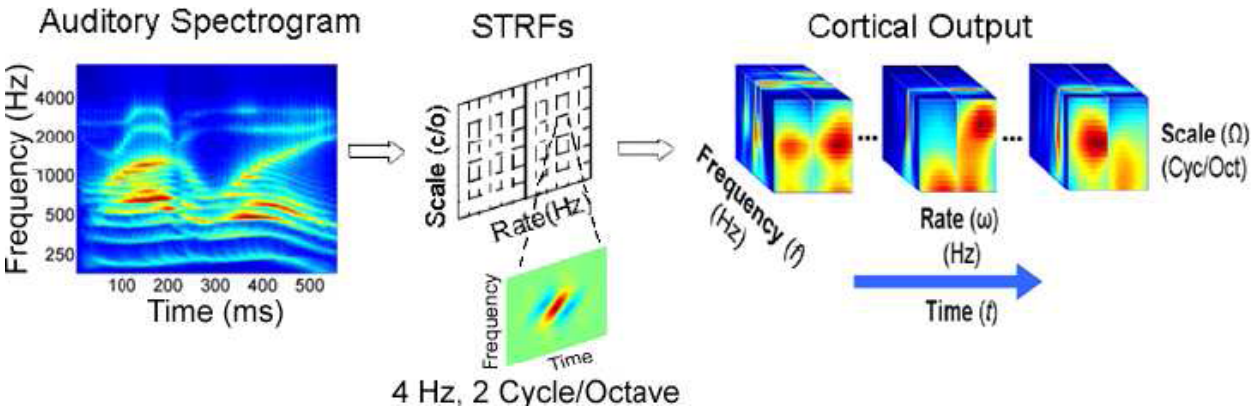
Επομένως, η φασματική-χρονική απόκριση ενός δεδομένου κυττάρου c του ακουστικού φλοιού για ένα φασματογράφημα εισόδου $y(t,f)$ θα είναι:

$$r_{c_+}(t,f;\omega_c,\Omega_c,\theta_c,\varphi_c) = y(t,f) * *_{t,f} STRF_+(t,f,\omega_c,\Omega_c,\theta_c,\varphi_c) \quad (2.15)$$

$$r_{c_-}(t,f,\omega_c,\Omega_c,\theta_c,\varphi_c) = y(t,f) * *_{t,f} STRF_-(t,f,\omega_c,\Omega_c,\theta_c,\varphi_c) \quad (2.16)$$

όπου με $**_{t,f}$ συμβολίζεται η δισδιάστατη συνέλιξη ως προς το χρόνο και τη συχνότητα.

Κατα την υλοποιήση του ακουστικού μοντέλου οι τιμές των παραμέτρων που χρησιμοποιήθηκαν βασίζονται σε ψυχοφυσιολογικά και ψυχοακουστικά ευρήματα [68, 35, 41]. Πιο συγκεκριμένα χρησιμοποιούνται οι θετικές $[+]$ και οι αρνητικές$[-]$ τιμές $rates \in \{2,4,8,16,32\}$



Σχήμα 2.4: Ηχητική αναπαράσταση φλοιού, (από το [106]).

(Hz) καθώς επισης χρησιμοποιούνται οι τιμές $scales \in \{0.25, 0.5, 1, 2, 4, 8\}$ (Cycles / Octave). Για κάθε ηχογράφηση εξάγεται, μια τετραδιάστατη αναπαράσταση φλοιού. Υπολογίζοντας τη μέση τιμή ως προς το χρόνο, προκύπτει μια τρισδιάστατη αναπαράσταση φλοιού η οποία αναπαρίσταται από έναν τανυστή τρίτης τάξης. Έστω, λοιπόν, ο τανυστής $\mathcal{D} \in \mathbb{R}_{+}^{I_1 \times I_2 \times I_3}$, όπου $I_1 = I_{scales} = 6$, $I_2 = I_{rates} = 10$, και $I_3 = I_{frequencies} = 128$ καλύπτοντας 128 συχνοτικά κανάλια κατά μήκος του τονοτοπικού άξονα.

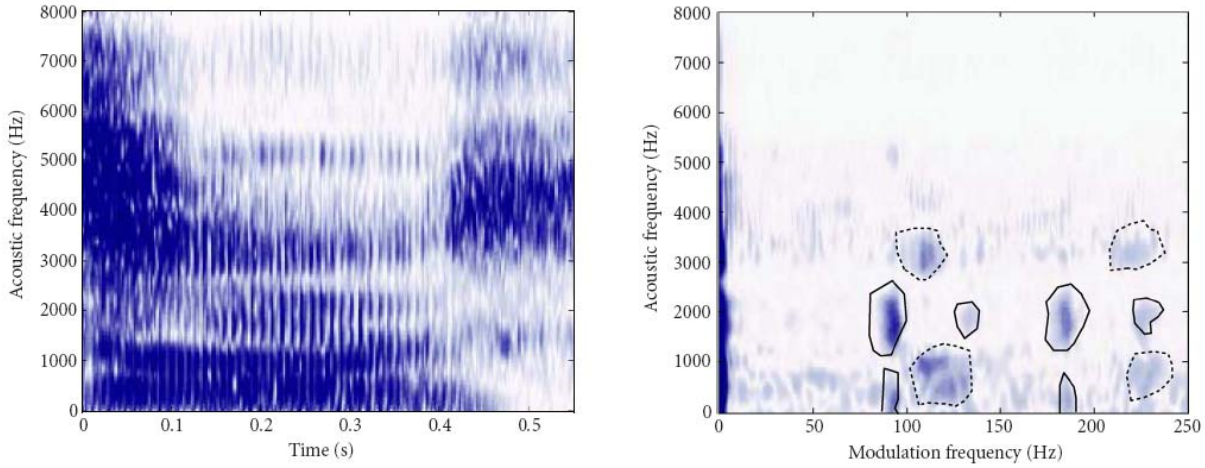
Οι εξισώσεις (2.15), (2.16) περιγράφουν μαθηματικά την *απο κοινού χρονοφασματική αναπαράσταση διαμορφώσεων του ακουστικού φλοιού* (cortical representation) του ήχου. Η εξαγωγή της αναπαράστασης φλοιού συνοψίζεται στο σχήμα 2.4. Για την ανωτερω διαδικασία χρησιμοποιήθηκε το NSL Matlab toolbox.

Το μέτρο των συναρτήσεων (2.15), (2.16) μπορεί να χρησιμοποιηθεί για την περεταίρω εξαγωγή χαρακτηριστικών. Πρόσφατα οι ηχητικές αναπαραστάσεις φλοιού χρησιμοποιήθηκαν με επιτυχία σε διάφορα προβλήματα ταξινόμησης ήχου με βάσει το περιεχόμενο. Οι Mesgarani [88], Wohlmayr [132] και Rifkin [106] χρησιμοποίησαν τις αναπαραστάσεις φλοιού με σκοπό να διακρίνουν σήματα ομιλίας από σήματα μη-ομιλίας ενώ ο Sundaram [110] για να διακρίνει δυο διαφορετικούς τύπους θορύβου.

2.4 Από κοινού συχνοτική ανάλυση

Η *απο κοινού συχνοτική ανάλυση* (joint frequency analysis) ή *φασματική ανάλυση διαμόρφωσης* έχει ως στόχο να μετασχηματίσει το ηχητικό σήμα, σε μια δομή που να αναπαριστά τις λανθάνουσες χρονικές διαμορφώσεις του σήματος. Η δομή αυτή είναι γνωστή ως *από κοινού συχνοτική αναπαράσταση* (joint frequency representation) [124].

Φορμαλιστικά, η από κοινού συχνοτική αναπαράσταση, έστω $P_s^{JF}(\omega, f)$ είναι ένας μετασχηματισμός στο χρόνο μιας διαμορφωμένης βραχυχρόνιας φασματικής εκτίμησης. Με ω η



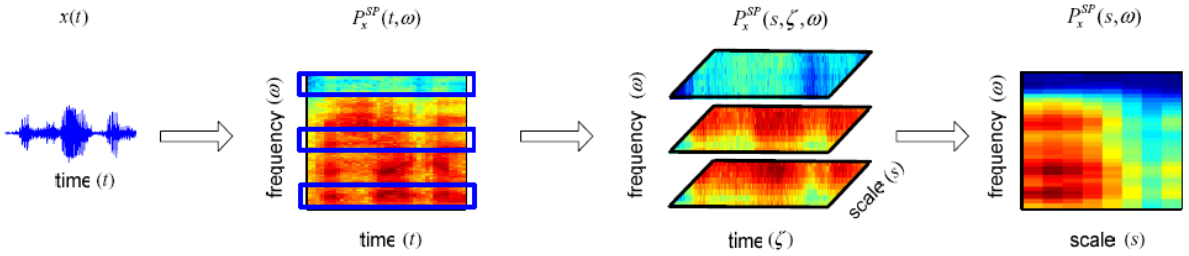
Σχήμα 2.5: Από κοινού συχνοτική αναπαράσταση, από το [2].

συχνότητα διαμόρφωσης ενώ με f υποδηλώνεται η *ακουστική συχνότητα*. Ένας τέτοιος μετασχηματισμός του σήματος $s(t)$ προκύπτει ως εξής : Αρχικά υπολογίζεται μια από κοινού χρονοσυχνοτική αναπαράσταση π.χ. το φασματογράφημα ή το ακουστικό φασματογράφημα, έστω $P_s^{SP}(t, f)$. Εν συνεχεία ένας δεύτερος μετασχηματισμός π.χ. (Fourier) ή συνημίτονου ή κυματιδίου εφαρμόζεται για κάθε κανάλι ακουστικής συχνότητας στον άξονα του χρόνου, ώστε να εκτιμηθεί η από κοινού συχνοτική αναπαράσταση του σήματος $P_s^{JF}(\omega, f)$ (βλ. σχήμα2.4).

Υπό μια διαφορετική σκοπιά ο μετασχηματισμός της από κοινού συχνοτικής ανάλυσης μπορεί να υπολογιστεί από την (2.17) δια της συνέλιξης ως προς ω των συναρτήσεων αυτοσυσχέτισης του μετασχηματισμού Fourier του σήματος $s(t)$, έστω $S(\omega)$, και του χρονικού παράθυρου ανάλυσης $w(t)$, έστω $W(\omega)$ [124, 125].

$$P_s^{JF}(\omega, f) = (W^*(\omega - \frac{f}{2})W(\omega + \frac{f}{2}) *_\omega (S^*(\omega - \frac{f}{2})S(\omega + \frac{f}{2}) \tag{2.17}$$

Η φασματική ανάλυση διαμόρφωσης έχει τη δυνατότητα να εξαγάγει τις χρονικά μεταβαλλόμενες πληροφορίες μέσω των μη μηδενικών όρων της από κοινού συχνοτικής αναπαράστασης. Όταν η ανάλυση εφαρμόζεται σε πραγματικό σήμα, π.χ. σήμα ομιλίας ή μουσικής, αυτοί οι μη-μηδενικοί όροι μπορούν να αντιπροσωπεύουν διάφορες φυσικές ποσότητες όπως φωνητικές πληροφορίες, την περίοδο μουσικού τόνου, ή τον ρυθμό και ενδεχομένως να περιέχουν χρήσιμη πληροφορία κατάλληλη για την ταξινόμηση των σημάτων σε κλάσεις. Εντούτοις, η χρήση της φασματικής ανάλυσης διαμόρφωσης που περιγράφεται από τη σχέση (2.17), δεν ενδείκνυται για την εξαγωγή χαρακτηριστικών γνωρισμάτων, διότι έχει το σοβαρό μειονέκτημα ότι παράγει χαρακτηριστικά γνωρίσματα πολύ μεγάλων διαστάσεων. Επιπλέον η χρήση της ανάλυσης Fourier, ή άλλων ομοιόμορφων μετασχηματισμών συχνότητας π.χ. μετασχηματισμός συνημίτονου, για τον υπολογισμό της συχνότητας διαμόρφωσης οδηγεί σε ένα μετασχηματισμό με ομοιόμορφα



Σχήμα 2.6: Modulation Scale Analysis, από το [125].

κατανεμημένο εύρος ζώνης συχνότητας στη διάσταση της συχνότητας διαμόρφωσης, γεγονός που δεν προσιδιάζει στην τονοτοπική οργάνωση των κυττάρων του πρωτοταγούς ακουστικού φλοιού όπως αυτή περιγράφηκε στην ενότητα 2.1.2. Ο Sukittanon σε μια προσπάθεια να μιμηθεί την απόκριση του ανθρώπινου ακουστικού συστήματος στις χρονικές διαμορφώσεις προτείνει την *κλιμακωτή ανάλυση διαμορφώσεων* (Modulation Scale Analysis) [125]. Η απόκριση των κυττάρων του κοχλία μπορεί να μοντελοποιηθεί από μια τράπεζα φίλτρων σταθερού Q. Ο μετασχηματισμός Q δύναται να προσεγγιστεί από τον μετασχηματισμό κυματιδίου συνεχούς χρόνου (continuoustime wavelet transform - CWT).

Το *φάσμα κλιμακωτής ανάλυσης διαμορφώσεων* (Modulation Scale Analysis Spectrum) αποτελεί την από κοινού αναπαράσταση της συχνότητας Fourier και της συχνότητας διαμόρφωσης με το ανομοιόμορφο εύρος ζώνης για την συχνότητα διαμόρφωσης. Όπως αναπαρίσταται στο Σχήμα 2.4, το φάσμα κλιμακωτής ανάλυσης διαμορφώσεων προκύπτει από μια ακολουθία τρίων βημάτων Η μαθηματική διατύπωση της κλιμακωτής ανάλυσης διαμορφώσεων έχει ως εξής: Αρχικά υπολογίζεται το φασματογράφημα του $s(t)$:

$$P_s^{SP}(t,f) = \frac{1}{2\pi} \mid \int s(u)w(u-t)e^{-j2\pi fu}du \mid^2 \tag{2.18}$$

Εν συνεχεία, εφαρμόζεται μετασχηματισμός κυματιδίου με βάση την συνάρτηση $\Psi(t)$ σε κάθε χρονική γραμμή του φασματογραφήματος για κάθε κλίμακα (scale) χρονικής διαμόρφωσης s.

$$P_s^{SP}(s,t,f) = \frac{1}{s} P_s^{SP}(t,f) *_t \Psi(-\frac{t}{s}) \tag{2.19}$$

Η ανωτέρω εξίσωση (2.19) μπορεί να αντιμετωπισθεί ως η εφαρμογή μετασχηματισμού κυματιδίου στη χρονική περιβάλλουσα κάθε ακουστικού καναλιού. Στο τελευταίο βήμα, η ενέργεια κατά τον άξονα του χρόνου t ολοκληρώνεται και έτσι προκύπτει το φάσμα κλιμακωτής ανάλυσης διαμορφώσεων:

$$P_s^{JF}(s,f) = \int \mid P_s^{SP}(s,t,f) \mid^2 dt \tag{2.20}$$

Υπάρχουν πολλά πλεονεκτήματα της παραπάνω μεθόδου από κοινού συχνοτική ανάλυση εν συγκρίσει με την από κοινού συχνοτικής ανάλυσης η οποία βασίζεται στον μετασχηματισμό Fourier. Η από κοινού συχνοτική ανάλυση που βασίζεται στον μετασχηματισμό κυματιδίου καταρχάς προσιδιάζει στο τρόπο που λειτουργεί το ανθρώπινο σύστημα ακοής, μπορεί να διαχωρίσει πολλαπλές συνιστώσες AM του σήματος και παράγει αναπαραστάσεις λίγων διαστάσεων.

Η απο κοινού συχνοτική ανάλυση έχει χρησιμοποιηθέι με επιτυχία ως μέθοδος εξαγωγής χαρακτηριστικών σε προβλήματα όπως αναγνώριση φωνής, ηχιτικών αποτυπωμάτων (audio fingerprinting) και κωδικοποίησης [124, 125, 2].

Στην παρούσα διατριβή προτείνουμε την αντικατάσταση του απλού φασματογραφήματος με το ακουστικό φασματογράφημα στην αλληλουχία υπολογισμού της δισδιάστατης από κοινού συχνοτικής αναπαράστασης σε μια προσπάθεια καλύτερης προσέγγισης του τρόπου με τον οποίο κωδικοποιείται η πληροφορία χρονικών διαμορφώσεων στον ακουστικό φλοιό. Το κοχλιακό μοντέλο που χρησιμοποιείται για τον υπολογισμό του ακουστικού φασματογραφήματος είναι αυτό του Lyon [83, 120], που ουσιαστικά πρόκειται για μια μικρή παραλλαγή αυτού που περιγράφεται στην ενότητα 2.3.1. Κατά την εξαγωγή του φάσματος κλιμακωτής ανάλυσης διαμορφώσεων οι τιμές των παραμέτρων που χρησιμοποιήθηκαν είναι για τις κλίμακες $s \in \{1, 2, 4, 8, 16, 32, 64, 128\}$ (Hz) ενώ το μοντέλο του κοχλία περιλαμβάνει 96 φίλτρα, 24 ανα οκτάβα, καλύπτοντας 4 οκτάβες κατα μήκος του τονοτοπικού άξονα. Για κάθε ηχογράφηση εξάγεται, μια τρισδίαστατη αναπαράσταση φλοιού χρονικών διαμορφώσεων από την (ΩT). Υπολογίζοντας τη μέση τιμή ως προς το χρόνο, προκύπτει μια δισδιάστατη αναπαράσταση φλοιού χρονικών διαμορφώσεων η οποία αναπαρίσταται από έναν τανυστή δέυτερης τάξης. Έστω, λοιπόν, ο τανυστής $\mathcal{D} \in \mathbb{R}_{+}^{I_1 \times I_2}$, όπου $I_1 = I_s = 8$, $I_2 = I_{frequencies} = 96$ καλύπτοντας 96 συχνοτικά κανάλια κατά μήκος του τονοτοπικού άξονα.

Κεφάλαιο 3

Λογισμός Τανυστών

Το παρόν κεφάλαιο περιέχει βασικούς ορισμούς και ιδιότητες της πολυγραμμικής άλγεβρας. Δίνεται ο ορισμός και οι βασικές ιδιότητες των τανυστών. Το κεφάλαιο αυτό διορθώνει, συμπληρώνει και επικαιροποιεί το αντίστοιχο κεφάλαιο στο [8].

3.1 Διανύσματα και Πίνακες

3.1.1 Ορισμοί

Ορισμός 3.1.1. (**Διανυσματικός Χώρος**) Διανυσματικός χώρος V πάνω σε ένα πεδίο F είναι ένα σύνολο στοιχείων, που λέγονται διανύσματα, $\mathbf{a}, \mathbf{b} \in V$ τέτοια ώστε για $\mathrm{c}, \mathrm{d} \in F$ να ισχύουν οι εξής ιδιότητες:

1. $(c \cdot \boldsymbol{\alpha} + d \cdot \boldsymbol{\beta}) \in V$
2. $\exists - \mathbf{a} \in V$: $\mathbf{a}$+(-$\mathbf{a}$)=$\mathbf{a}$-$\mathbf{a}$=0
3. $\exists o \in V : 0 + \mathbf{a} = \mathbf{a}$
4. $1 \cdot \mathbf{a} = \mathbf{a},\ \forall \mathbf{a} \in V$
5. $\mathbf{a}$+$\mathbf{b}$=$\mathbf{b}$+$\mathbf{a}$
6. $\mathbf{a}$+($\mathbf{b}$+$\mathbf{c}$)=($\mathbf{a}$+$\mathbf{b}$)+$\mathbf{c}$
7. $(c + d) \cdot \mathbf{a} = c \cdot \mathbf{a} + d \cdot \mathbf{a}$,
   $c \cdot (\mathbf{a} + \mathbf{b}) = c \cdot \mathbf{a} + c \cdot \mathbf{b}$,
   $(cd) \cdot \mathbf{a} = c \cdot (d\mathbf{a})$.

□

Παράδειγμα 3.1.1. *Ένα* $\mathbf{a}$ *διάνυσμα γραμμή μήκους* n *συμβολίζεται ως:*

$$\mathbf{a} = \begin{pmatrix} a_1 & a_2 & \dots & a_n \end{pmatrix}. \tag{3.1}$$

□

Ορισμός 3.1.2. (**Πίνακας**) Έστω V_1, V_2 διανυσματικοί χώροι με διαστάσεις, I_1, I_2, αντίστοιχα. Θεωρούμε 2 διανύσματα $\mathbf{u}_1 \in V_1, \mathbf{u}_2 \in V_2$. Ο χώρος που δημιουργείται από όλα τα στοιχεία της διγραμμικής απεικόνισης $\mathbf{u}_1 \circ \mathbf{u}_2$ στον $V_1 \times V_2$ είναι ο χώρος διανυσματικού γινομένου για διανύσματα από τους χώρους V_1, V_2. Ένα στοιχείο του χώρου διανυσματικού γινομένου ονομάζεται πίνακας (ή μήτρα) στον $V_1 \times V_2$. □

Ένας πιο πρακτικός ορισμός του πίνακα είναι ο εξής: μία δισδιάστατη διάταξη στοιχείων ενός πεδίου F, που έχει m γραμμές και n στήλες, ονομάζεται πίνακας $m \times n$ στο F. Συμβολίζεται ως $\mathbf{A} = [a_{ij}]_{mn}$, όπου οι δείκτες mn ορίζουν τις δυναμικές περιοχές των δεικτών ij, δηλαδή $i = 1, \ldots, m$ και $j = 1, \ldots, n$. Συμφωνούμε ότι όλα τα διανύσματα θα εννοούνται ως διανύσματα γραμμής εφεξής, οπότε τα διανύσματα είναι $1 \times n$ πίνακες.

Παράδειγμα 3.1.2. *Έστω* $\mathbf{A}$ *πίνακας* $m \times n$ *διαστάσεων, τότε*

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix}. \tag{3.2}$$

□

3.1.2 Ιδιότητες

Ορισμός 3.1.3. (**Εσωτερικό Γινόμενο Διανυσμάτων**) Αν $\mathbf{u} = (u_1 \ldots u_n)$ και $\mathbf{v} = (v_1 \ldots v_n)$, το εσωτερικό γινόμενο των $\mathbf{u}$ και $\mathbf{v}$ ορίζεται ως:

$$\mathbf{u} \cdot \mathbf{v} = \sum_{i=1}^{n} u_i v_i. \tag{3.3}$$

Τα διανύσματα $\mathbf{u}$ και $\mathbf{v}$ είναι ορθογώνια, αν $\mathbf{u} \cdot \mathbf{v} = 0$. □

Ορισμός 3.1.4. (**Μοναδιαίος Πίνακας**) Ορίζεται ως ο τετραγωνικός πίνακας ($m = n$):

$$\mathbf{I} = [\delta_{ij}]_{nn}. \tag{3.4}$$

όπου δ_{ij} το δέλτα του Kronecker. Για οποιονδήποτε τετραγωνικό πίνακα $\mathbf{A}$ διαστάσεων $n \times n$ ισχύει ότι: $\mathbf{I} \cdot \mathbf{A} = \mathbf{A} \cdot \mathbf{I} = \mathbf{A}$. □

Ορισμός 3.1.5. (**Νόρμα Frobenius**) Για έναν πίνακα $\mathbf{A}$ διαστάσεων $m \times n$ ορίζεται ως:

$$\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2}. \tag{3.5}$$

□

Ορισμός 3.1.6. (**Γινόμενο Πινάκων**) Αν $\mathbf{A} = [a_{ij}]_{mn}$ και $\mathbf{B} = [b_{ij}]_{nk}$, το γινόμενο των πινάκων $\mathbf{A}$ και $\mathbf{B}$ συμβολίζεται $\mathbf{AB} = [(\mathbf{AB})_{ij}]_{mk}$, και έχει ij στοιχείο:

$$(\mathbf{AB})_{ij} = \sum_{r=1}^{n} a_{ir} b_{rj}. \tag{3.6}$$

□

Ορισμός 3.1.7. (**Ορθογώνιος Τετραγωνικός Πίνακας**) Ένας τετραγωνικός πίνακας $\mathbf{A} = [a_{ij}]_{nn}$ είναι ορθογώνιος αν:

$$\mathbf{A}^T = \mathbf{A}^{-1}. \tag{3.7}$$

όπου $\mathbf{A}^T = [a_{ji}]_{nn}$ είναι ο ανάστροφος του $\mathbf{A}$, ενώ $\mathbf{A}^{-1}$ είναι ο αντίστροφος του $\mathbf{A}$, δηλαδή πίνακας τέτοιος ώστε $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$. □

3.2 Ορισμός Τανυστών

Οι *τανυστές* (tensors) αποτελούν μία γενίκευση των διανυσμάτων και των πινάκων και αποτελούν βασικό αντικείμενο της *Πολυγραμμικής Άλγεβρας*. Ανάλογα με το επιστημονικό πεδίο στο οποίο συναντώνται, οι τανυστές μπορούν να οριστούν με διαφορετικούς τρόπους. Στο μαθηματικό πεδίο της *Ανάλυσης Τανυστών*, ορίζονται ως αντικείμενα που υπόκεινται σε πολυδιάστατους μετασχηματισμούς, από ένα σύστημα συντεταγμένων σε ένα άλλο. Οι μετασχηματισμοί που χρησιμοποιούνται ονομάζονται συμμεταβλητοί (covariant - από τον διανυσματικό στο δυϊκό χώρο) και αντιμεταβλητοί (contravariant - από τον δυϊκό χώρο στο διανυσματικό). Στη φυσική χρησιμοποιούνται για να περιγράψουν ποσότητες στον τρισδιάστατο χώρο. Οι τανυστές χρησιμοποιούνται επίσης για την περιγραφή πεδίων. Για τους παραπάνω λόγους, χρησιμοποιούνται εκτενώς στο πεδίο της μηχανικής συνεχών μέσων (continuum mechanics). Περισσότερες πληροφορίες για το πεδίο της πολυγραμμικής άλγεβρας μπορούν να αναζητηθούν στο [49].

Στον κλάδο των εφαρμοσμένων μαθηματικών, όπου δεν απαιτούνται μετασχηματισμοί σε συστήματα συντεταγμένων, οι τανυστές ορίζονται ως μία γενικευμένη γραμμική οντότητα που μπορεί να αναπαρασταθεί ως ένας πολλαπλός (multiway) πίνακας. Ακολουθεί ο ορισμός των τανυστών που μπορεί να εφαρμοστεί μόνο για γραμμικούς μετασχηματισμούς συντεταγμένων (component-free approach) [72]. Για τη χρήση τανυστών σε εφαρμογές που χρησιμοποιούνται ως πολλαπλοί πίνακες, έχει προταθεί το Tensor Toolbox για περιβάλλον MATLAB[1] καθώς επίσης και το N-way Matlab Toolbox[2].

[1] cmr.ca.sandia.gov/ tgkolda/TensorToolbox/

[2] www.models.klv.dk/source/nwaytoolbox

Ορισμός 3.2.1. (**Τανυστής N-οστής Τάξης**) Έστω $V_1, V_2, \ldots, V_N$ N διανυσματικοί χώροι με διαστάσεις $I_1, I_2, \ldots, I_N$. Θεωρούμε N διανύσματα $\mathbf{u}_1 \in V_1, \mathbf{u}_2 \in V_2, \ldots, \mathbf{u}_N \in V_N$. Ο χώρος που δημιουργείται από όλα τα στοιχεία της πολυγραμμικής απεικόνισης $\mathbf{u}_1 \circ \mathbf{u}_2 \circ \ldots \circ \mathbf{u}_N$ στον $V_1 \times V_2 \times \ldots \times V_N$ ονομάζεται χώρος τανυστών. Ένα στοιχείο του χώρου τανυστών ονομάζεται τανυστής N-οστής τάξης στον $V_1 \times V_2 \times \ldots \times V_N$. □

Για $V_n = \mathbb{R}^{I_n}$ $(n = 1, 2, \ldots, N)$, ο χώρος γινομένου τανυστών ονομάζεται χώρος των πραγματικών $(I_1 \times I_2 \times \ldots \times I_N)$-τανυστών, και συμβολίζεται με $\mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$. Ο αντίστοιχος μιγαδικός χώρος συμβολίζεται με $\mathbb{C}^{I_1 \times I_2 \times \ldots \times I_N}$.

Παράδειγμα 3.2.1. *Έστω τανυστής $\mathcal{A} \in \mathbb{R}^{3\times3\times3}$. Θεωρούμε ότι $a_{ij1} = 1, a_{ij2} = 2, a_{ij3} = 3$ $(i, j = 1, 2, 3)$. Ο $\mathcal{A}$ είναι δυνατόν να συμβολιστεί χρησιμοποιώντας πίνακες:*

$$\mathcal{A} = [A_{ij1}|A_{ij2}|A_{ij1}] \ \ \mu\epsilon \ \ A_{ij1} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \ \ A_{ij2} = \begin{pmatrix} 2 & 2 & 2 \\ 2 & 2 & 2 \\ 2 & 2 & 2 \end{pmatrix} \ \ A_{ij3} = \begin{pmatrix} 3 & 3 & 3 \\ 3 & 3 & 3 \\ 3 & 3 & 3 \end{pmatrix}. \quad (3.8)$$

□

Ορισμός των μικτών τανυστών που υπόκεινται σε μη γραμμικούς μετασχηματισμούς και απαντιούνται σε καμπυλόγραμμες συντεταγμένες παρατίθεται στα [15, 65]. Η ανάλυσή μας εμπλουτίζεται όμως με σχόλια σχετικά με τη συμπεριφορά των μικτών τανυστών για λόγους σύγκρισης. Για λόγους πληρότητας παρατίθεται ο ορισμός των τανυστών με χρήση μη-γραμμικών μετασχηματισμών (component approach):

Ορισμός 3.2.2. (**Μικτός Τανυστής N-οστής Τάξης**) Ένας τανυστής $\mathcal{A}$ ορίζεται ως μικτός τανυστής τάξης $(P + Q)$, αντιμεταβλητής τάξης P και συμμεταβλητής τάξης Q, αν τα στοιχεία του $\alpha_{i_1 i_2 \ldots i_P}^{j_1 j_2 \ldots j_Q}$ μετασχηματίζονται σε $\tilde{\alpha}_{i_1 i_2 \ldots i_P}^{j_1 j_2 \ldots j_Q}$ μεταβαίνοντας από το σύστημα συντεταγμένων $x_{i_p}^{(p)}$ στο σύστημα συντεταγμένων $\tilde{x}_{i_p}^{(p)}$ για $p = 1, \ldots, P$ και από $y_{j_q}^{(q)}$ σε $\tilde{y}_{j_q}^{(q)}$ για $q = 1, \ldots, Q$ με τον ακόλουθο κανόνα:

$$\tilde{\alpha}_{i_1 i_2 \ldots i_P}^{j_1 j_2 \ldots j_Q} = \alpha_{r_1 r_2 \ldots r_P}^{s_1 s_2 \ldots s_Q} \frac{\partial y_{s_1}^{(1)}}{\partial \tilde{y}_{j_1}^{(1)}} \frac{\partial y_{s_2}^{(2)}}{\partial \tilde{y}_{j_2}^{(2)}} \cdots \frac{\partial y_{s_Q}^{(Q)}}{\partial \tilde{y}_{j_Q}^{(Q)}} \frac{\partial x_{i_1}^{(1)}}{\partial \tilde{x}_{r_1}^{(1)}} \frac{\partial x_{i_2}^{(2)}}{\partial \tilde{x}_{r_2}^{(2)}} \cdots \frac{\partial x_{i_P}^{(P)}}{\partial \tilde{x}_{r_P}^{(P)}}. \quad (3.9)$$

Η εξίσωση (3.9) είναι γνωστή ως *Κανόνας Μετασχηματισμού Τανυστών.* □

3.3 Πράξεις με Τανυστές

Για τανυστές πεπερασμένης τάξης και διαστάσεων ορίζονται και οι βασικές πράξεις της πολυγραμμικής άλγεβρας, που αποτελούν γενίκευση των βασικών πράξεων της γραμμικής άλγεβρας. Οι ορισμοί δίνονται για μιγαδικούς τανυστές, και προφανώς μπορούν να εφαρμοστούν και σε τανυστές με πραγματικές τιμές.

3.3.1 Βασικές Πράξεις

Ορισμός 3.3.1. (**Πρόσθεση Τανυστών**) Έστω δύο τανυστές $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \ldots \times I_N}$ και $\mathcal{B} \in \mathbb{C}^{I_1 \times I_2 \times \ldots \times I_N}$. Το άθροισμά τους $(\mathcal{A} + \mathcal{B}) \in \mathbb{C}^{I_1 \times I_2 \times \ldots \times I_N}$ είναι ο τανυστής με στοιχεία:

$$(\mathcal{A} + \mathcal{B})_{i_1 i_2 \ldots i_N} = a_{i_1 i_2 \ldots i_N} + b_{i_1 i_2 \ldots i_N}. \tag{3.10}$$

□

Ανάλογα ορίζεται και το βαθμωτό γινόμενο ενός τανυστή $\mathcal{A}$ με μία πραγματική σταθερά λ, που δημιουργεί έναν τανυστή ίδιας τάξης και διαστάσεων με τον $\mathcal{A}$.

Ορισμός 3.3.2. (**Γραμμικός Συνδυασμός**) Θεωρούμε M τανυστές $\mathcal{A}^{(l)} \in \mathbb{C}^{I_1 \times I_2 \times \ldots \times I_N}$, $l = 1, 2, \ldots, M$, και M σταθερές $\lambda_1, \lambda_2, \ldots, \lambda_M \in \mathbb{R}$. Ο τανυστής $\mathcal{B} = \lambda_1 \mathcal{A}^{(1)} + \lambda_2 \mathcal{A}^{(2)} + \ldots + \lambda_M \mathcal{A}^{(M)}$ έχει την ίδια τάξη και διαστάσεις με τους $\mathcal{A}^{(l)}$ $(l = 1, 2, \ldots, M)$. □

3.3.2 Γινόμενα Τανυστών

Ορισμός 3.3.3. (**Εσωτερικό Γινόμενο**) Έστω τανυστές $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \ldots \times I_M}$ και $\mathcal{B} \in \mathbb{C}^{J_1 \times J_2 \times \ldots \times J_N}$. Το εσωτερικό γινόμενο ως προς τους δείκτες i_m και i_n, με κοινή διάσταση $|I_m| = |J_n| = p$, ορίζεται ως $\langle \mathcal{A}, \mathcal{B} \rangle_{m,n} \in \mathbb{C}^{I_1 \times I_2 \times \ldots \times I_{m-1} \times I_{m+1} \times \ldots I_M \times J_1 \times J_2 \times \ldots \times J_{n-1} \times J_{n+1} \times \ldots \times J_N}$:

$$(\langle \mathcal{A}, \mathcal{B} \rangle_{m,n})_{i_1 \ldots i_{m-1} i_{m+1} \ldots i_M j_1 \ldots j_{n-1} j_{n+1} \ldots j_N} = \sum_{k=1}^{p} a_{i_1 \ldots i_{m-1} k i_{m+1} \ldots i_M} b_{j_1 \ldots j_{n-1} k j_{n+1} \ldots j_N}. \tag{3.11}$$

□

Είναι δυνατόν να οριστεί εσωτερικό γινόμενο πάνω σε πολλούς δείκτες. Επίσης, στην περίπτωση μικτού τανυστή, το εσωτερικό γινόμενο ορίζεται μόνο ανάμεσα σε συμμεταβλητούς και αντιμεταβλητούς δείκτες. Πρακτικά, οι συμμεταβλητές και αντιμεταβλητές συμπεριφορές του τανυστή αλληλοακυρώνονται, μειώνοντας την συνολική τάξη του τανυστή.

Στην περίπτωση τανυστών 2ης τάξης (πίνακες), το εσωτερικό γινόμενο ισούται με το γινόμενο πινάκων (Ορισμός 3.1.6). Το αποτέλεσμα είναι ένας τανυστής 2ης τάξης ο οποίος ορίζεται στις δύο μη κοινές διαστάσεις των πινάκων.

Έστω τανυστές $\mathcal{A}, \mathcal{B} \in \mathbb{R}^{2 \times 2 \times 2}$. Θεωρούμε ότι $a_{i_1 i_2 i_3} = 1, b_{j_1 j_2 j_3} = 2$ $(1 \leqslant i_x, j_x \leqslant 2)$ $x = 1, 2, 3$. Υπολογίζουμε το εσωτερικό γινόμενο των $\mathcal{A}, \mathcal{B}$ με βάση την πρώτη διάσταση ως κοινή:

$$(\langle \mathcal{A}, \mathcal{B} \rangle_{1,1})_{i_2 i_3 j_2 j_3} = \sum_{k=1}^{2} a_{k i_2 i_3} b_{k j_2 j_3}. \tag{3.12}$$

Για να υπολογίσουμε τα στοιχεία του τανυστή $\langle \mathcal{A}, \mathcal{B} \rangle_{1,1} \in \mathbb{R}^{2 \times 2 \times 2 \times 2}$ χρησιμοποιούμε την (3.12), π.χ. το στοιχείο (1,1,1,1): $(\langle \mathcal{A}, \mathcal{B} \rangle_{1,1})_{1111} = a_{111} b_{111} + a_{211} b_{211} = 2 + 2 = 4$. □

Ορισμός 3.3.4. (**Εξωτερικό Γινόμενο**) Έστω τανυστές $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \ldots \times I_M}$ και $\mathcal{B} \in \mathbb{C}^{J_1 \times J_2 \times \ldots \times J_N}$. Το εξωτερικό γινόμενο $\mathcal{A} \otimes \mathcal{B} \in \mathbb{C}^{I_1 \times \ldots \times I_M \times J_1 \times \ldots \times J_N}$, των τανυστών $\mathcal{A}$ και $\mathcal{B}$ είναι ο τανυστής με στοιχεία:

$$(\mathcal{A} \otimes \mathcal{B})_{i_1 \ldots i_M j_1 \ldots j_N} = a_{i_1 \ldots i_M} b_{j_1 \ldots j_N}. \tag{3.13}$$

□

Άρα το εξωτερικό γινόμενο δύο τανυστών (το οποίο συναντάται στην βιβλιογραφία και ως τανυστικό γινόμενο) έχει τάξη $M + N$, και διαστάσεις $I_1 \times \ldots \times I_M \times J_1 \times \ldots \times J_N$. Στην περίπτωση πινάκων (τανυστών τάξης 2) στον ίδιο διανυσματικό χώρο, το εξωτερικό γινόμενο ονομάζεται γινόμενο Kronecker. Στην περίπτωση εξωτερικού γινομένου μικτών τανυστών, προκύπτει τανυστής με συμμεταβλητή τάξη ίση με το άθροισμα της συμμεταβλητής τάξης των δύο και αντίστοιχα προκύπτει τανυστής με αντιμεταβλητή τάξη ίση με το άθροισμα της αντιμεταβλητής τάξης των δύο.

Έστω πίνακες $\mathbf{A} \in \mathbb{R}^{I_1 \times I_2}$ διαστάσεων 2×2 με $a_{ij} = 1$ και $\mathbf{B} \in \mathbb{R}^{I_3 \times I_4}$ ίδιων διαστάσεων με $b_{kl} = 2$. Το εξωτερικό γινόμενο των δύο πινάκων είναι τανυστής με στοιχεία:

$$(\mathbf{A} \otimes \mathbf{B})_{i_1 i_2 i_3 i_4} = a_{i_1 i_2} b_{i_3 i_4}. \tag{3.14}$$

Άρα το προκύπτον εξωτερικό γινόμενο είναι τανυστής 4ης τάξης διαστάσεων $2 \times 2 \times 2 \times 2$. Το στοιχείο (1,1,1,1) του εξωτερικού γινομένου ισούται με: $(\mathbf{A} \otimes \mathbf{B})_{1234} = 1 \cdot 2 = 2$. Στην περίπτωση που $\mathbf{B} \in \mathbb{R}^{I_1 \times I_2}$, δηλαδή οι πίνακες $\mathbf{A}$ και $\mathbf{B}$ ανήκουν στον ίδιο διανυσματικό χώρο, τότε το εξωτερικό τους γινόμενο ισούται με το γινόμενο Kronecker:

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} 2 & 2 & 2 & 2 \\ 2 & 2 & 2 & 2 \\ 2 & 2 & 2 & 2 \\ 2 & 2 & 2 & 2 \end{pmatrix}. \tag{3.15}$$

Σ′ αυτήν την περίπτωση το αποτέλεσμα είναι πίνακας διαστάσεων $2 \cdot 2 \times 2 \cdot 2 = 4 \times 4$. □

3.3.3 Μοναδιαίος και Ισοτροπικός Τανυστής

Ορισμός 3.3.5. (**Μοναδιαίος Τανυστής**) Ο μοναδιαίος τανυστής $\mathcal{I} \in \mathbb{C}^{I_1 \times \ldots \times I_N \times J_1 \times \ldots \times J_N}$ ορίζεται ως μία πολυμεταβλητή γενίκευση του δέλτα του Kronecker και έχει στοιχεία:

$$\mathcal{I}_{i_1 \ldots i_N j_1 \ldots j_N} = \delta_{i_1 \ldots i_N}^{j_1 \ldots j_N} = \prod_{k=1}^{N} \delta_{i_k j_k}. \tag{3.16}$$

□

Πρακτικά, ο μοναδιαίος τανυστής κατασκευάζεται από το εξωτερικό γινόμενο πολλών μοναδιαίων πινάκων $\mathbf{I}$, και ορίζεται για άρτιες τάξεις όπως επεξηγείται στο επόμενο παράδειγμα.

Παράδειγμα 3.3.1. *Ο μοναδιαίος τανυστής* $\mathcal{I} \in \mathbb{C}^{I_1 \times I_2 \times J_1 \times J_2}$ *έχει στοιχεία:*

$$\mathcal{I}_{i_1 i_2 j_1 j_2} = \delta_{i_1 j_1} \delta_{i_2 j_2}. \tag{3.17}$$

δηλαδή έχει τιμή 1 όταν $i_1 = j_1$ *και* $i_2 = j_2$ *και 0 αλλού. Αν θεωρήσουμε ότι ο* $\mathcal{I}$ *έχει διαστάσεις* $2 \times 2 \times 2 \times 2$*, τότε παίρνει τιμή 1 στα στοιχεία: (1,1,1,1), (1,2,1,2), (2,1,2,1), (2,2,2,2) ενώ παίρνει τιμή 0 στα στοιχεία (1,1,1,2), (1,1,2,1), (1,2,1,1), (2,1,1,1), (1,2,2,2), (1,2,2,1), (1,1,2,2), (2,2,1,1), (2,2,2,1), (2,1,1,2), (2,1,2,2), και (2,2,1,2).* □

Επειδή τα στοιχεία του μοναδιαίου τανυστή δεν εξαρτώνται από συγκεκριμένη επιλογή βάσης, ο μοναδιαίος τανυστής ονομάζεται *ισοτροπικός* (isotropic). Συγκεκριμένα, ένας τανυστής του οποίου οι τιμές των στοιχείων του είναι αναλλοίωτες με την περιστροφή του συστήματος συντεταγμένων ονομάζεται ισοτροπικός. Όλοι οι τανυστές μηδενικής τάξης (βαθμωτά μεγέθη) είναι ισοτροπικοί, ενώ δεν υπάρχουν ισοτροπικά διανύσματα. Ο μοναδικός ισοτροπικός τανυστής δεύτερης τάξης είναι το δέλτα του Kronecker. Ο μοναδικός ισοτροπικός τανυστής τρίτης τάξης είναι το *σύμβολο αντιμετάθεσης* ϵ_{ijk}:

$$\epsilon_{ijk} = \begin{cases} 0 & \text{για } i = j, \text{ ή } j = k, \text{ ή } k = i \\ 1 & \text{για } (i,j,k) \in \{(1,2,3),(2,3,1),(3,1,2)\} \\ -1 & \text{για } (i,j,k) \in \{(1,3,2),(3,2,1),(2,1,3)\} \end{cases} \tag{3.18}$$

ο οποίος στην βιβλιογραφία αναφέρεται ως *τανυστής Levi-Civita*.

3.3.4 Συστολή Τανυστή

Ορισμός 3.3.6. (**Συστολή Τανυστή**) Θεωρούμε τανυστή $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \ldots \times I_N}$. Η συστολή (contraction) του $\mathcal{A}$ ως προς τους δείκτες i_p και i_q, με κοινή διάσταση u, είναι ο τανυστής $\langle \mathcal{A} \rangle_{p,q}$ τάξεως $(N-2)$, με στοιχεία:

$$(\langle \mathcal{A} \rangle_{p,q})_{i_1 \ldots i_{p-1} i_{p+1} \ldots i_{q-1} i_{q+1} \ldots i_N} = \sum_{k=1}^{u} a_{i_1 \ldots i_{p-1} k i_{p+1} \ldots i_{q-1} k i_{q+1} \ldots i_N}. \tag{3.19}$$

□

Η συστολή ενός τανυστή μειώνει την τάξη του κατά 2. Είναι επίσης δυνατόν να οριστεί συστολή για περισσότερους από 2 δείκτες. Στην περίπτωση των πινάκων, η συστολή είναι αντίστοιχη με το ίχνος (trace) του πίνακα. Στην περίπτωση μικτού τανυστή, η συστολή ορίζεται ανάμεσα ως προς έναν συμμεταβλητό και έναν αντιμεταβλητό δείκτη.

Αξίζει να σημειωθεί η σχέση του εσωτερικού γινομένου, του εξωτερικού γινομένου και της συστολής τανυστών. Το εξωτερικό γινόμενο των τανυστών $\mathcal{A}$ και$\mathcal{B}$ και η ακολούθως συστολή του γινομένου ως προς δεδομένους δείκτες, δίνει ίδιο αποτέλεσμα με το εσωτερικό γινόμενο των δύο τανυστών ως προς τους ίδιους δείκτες.

Θεωρούμε τανυστή $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times I_3 \times I_4}$, με διαστάσεις $2 \times 2 \times 2 \times 2$ και τιμές $a_{i_1 i_2 i_3 i_4} = 1$. Η συστολή του $\mathcal{A}$ ως προς τους δείκτες i_2, i_3 είναι:

$$(\langle \mathcal{A} \rangle_{2,3})_{i_1,i_4} = \sum_{k=1}^{2} a_{i_1 k k i_4} = a_{i_1 11 i_4} + a_{i_1 22 i_4}. \tag{3.20}$$

Άρα το αποτέλεσμα είναι ο πίνακας $\mathbf{A} \in \mathbb{R}^{I_1 \times I_4}$ με στοιχεία:

$$(\langle \mathcal{A} \rangle_{2,3})_{i_1,i_4} = \mathbf{A} = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix}. \tag{3.21}$$

□

3.3.5 Γινόμενο Τανυστή με Πίνακα

Ορισμός 3.3.7. (**Γινόμενο Τανυστή με Πίνακα**) Θεωρούμε τανυστή $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \ldots \times I_N}$ και πίνακα $\mathbf{B} \in \mathbb{C}^{J_n \times I_n}$. Το γινόμενο του $\mathcal{A}$ επί τον $\mathbf{B}$ ως προς την κοινή διάσταση I_n ορίζεται ως:

$$\mathcal{A} \times_n \mathbf{B} = \langle \mathcal{A}, \mathbf{B} \rangle_{n,2}. \tag{3.22}$$

□

Κοινώς, το γινόμενο τανυστή με πίνακα ορίζεται ως το εσωτερικό γινόμενό τους πάνω σε κοινούς δείκτες, και παράγει τανυστή τάξεως N όπου η δίασταση του n-οστού δείκτη μεταβάλλεται από I_n σε J_n και τα στοιχεία του γινομένου υπολογίζονται με βάση την παρακάτω σχέση:

$$(\mathcal{A} \times_n \mathbf{B})_{i_1 \ldots i_{n-1} j_n i_{n+1} \ldots i_N} = \sum_{k=1}^{I_n} a_{i_1 \ldots k \ldots i_N} b_{j_n k}. \tag{3.23}$$

Παρατηρούμε ότι το γινόμενο τανυστή με πίνακα, ο τανυστής που προκύπτει έχει την ίδια τάξη με τον αρχικό τανυστή: $(\mathcal{A} \times_n \mathbf{B}) \in \mathbb{C}^{I_1 \times \ldots \times J_n \times \ldots \times I_N}$. Απλά ο n-οστός δείκτης ανήκει πλέον στο J_n.

Αν θεωρήσουμε τον τανυστή $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \ldots \times I_N}$ και πίνακες $\mathbf{B} \in \mathbb{C}^{J_n \times I_n}$, $\mathbf{C} \in \mathbb{C}^{J_m \times I_m}$ και $\mathbb{D}^{K_n \times J_n}$ τότε ορίζονται δύο ιδιότητες που αφορούν το γινόμενο τανυστή με πίνακες:

$$(\mathcal{A} \times_n \mathbf{B}) \times_m \mathbf{C} = (\mathcal{A} \times_m \mathbf{C}) \times_n \mathbf{B} = \mathcal{A} \times_n \mathbf{B} \times_m \mathbf{C}. \tag{3.24}$$

Ας θεωρήσουμε τον πίνακα $\mathbf{C} \in \mathbb{C}^{K_n \times J_n}$, τότε

$$(\mathcal{A} \times_n \mathbf{B}) \times_n \mathbf{C} = \mathcal{A} \times_n (\mathbf{C} \cdot \mathbf{B}). \tag{3.25}$$

Η εξίσωση (3.24) είναι η προσεταιριστική ιδιότητα του πολλαπλασιασμού τανυστή με πίνακες. Το σύμβολο $\cdot$ στην (3.25) εκφράζει τον πολλαπλασιασμό δύο πινάκων, όπως διατυπώθηκε στον ορισμό (3.6).

Στην περίπτωση γινομένου 3 πινάκων, έστω $\mathbf{A} \in \mathbb{C}^{I_1 \times I_2}$, $\mathbf{B} \in \mathbb{C}^{I_3 \times I_1}$, και $\mathbf{C} \in \mathbb{C}^{I_4 \times I_2}$, το γινόμενο $\mathbf{B} \cdot \mathbf{A} \cdot \mathbf{C}^H$ μπορεί να γραφεί ως: $\mathbf{A} \times_1 \mathbf{B} \times_2 \mathbf{C}$. Φαίνεται καθαρά ότι ο $\mathbf{B}$ πολλαπλασιάζεται ως προς την I_1 διάσταση, ενώ ο $\mathbf{C}$ ως προς την I_2 διάσταση. Στον N-διάστατο χώρο ο παραπάνω συμβολισμός λύνει το πρόβλημα των συμβολισμών για γενικευμένους ανάστροφους πινάκων.

Θεωρούμε τανυστή τρίτης τάξεως $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ με διαστάσεις $2 \times 2 \times 2$ και τιμές $a_{i_1 i_2 i_3} = 1$. Επίσης θεωρούμε πίνακα $\mathbf{B} \in \mathbb{R}^{I_4 \times I_1}$ διαστάσεων 2×2:

$$\mathbf{B} = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}.$$

Το γινόμενο $\mathcal{A} \times_1 \mathbf{B}$ δίνεται από την παρακάτω σχέση:

$$(\mathcal{A} \times_1 \mathbf{B})_{i_4 i_2 i_3} = \sum_{k=1}^{2} a_{k i_2 i_3} b_{i_4 k} = a_{1 i_2 i_3} b_{i_4 k} + a_{2 i_2 i_3} b_{i_4 k}.$$

Άρα το γινόμενο παίρνει τις εξής τιμές: $(\mathcal{A} \times_1 \mathbf{B})_{1 i_2 i_3} = 3$, $(\mathcal{A} \times_1 \mathbf{B})_{2 i_2 i_3} = 2$.

$$\mathcal{A} \times_1 \mathbf{B} = \left(\begin{array}{cc|cc} 3 & 3 & 2 & 2 \\ 3 & 3 & 2 & 2 \end{array} \right)$$

.

□

3.3.6 Βαθμωτό Γινόμενο

Ορισμός 3.3.8. (Βαθμωτό Γινόμενο Τανυστών) Θεωρούμε τανυστές $\mathcal{A}, \mathcal{B} \in \mathbb{C}^{I_1 \times I_2 \times \ldots \times I_N}$. Το βαθμωτό γινόμενο (scalar product) των $\mathcal{A}$ και $\mathcal{B}$ ορίζεται ως:

$$\langle \mathcal{A}, \mathcal{B} \rangle = \sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \ldots \sum_{i_N=1}^{I_N} a_{i_1 i_2 \ldots i_N} b^*_{i_1 i_2 \ldots i_N}. \tag{3.26}$$

□

Όπως φαίνεται από την (3.26), το βαθμωτό γινόμενο τανυστών αποτελεί γενίκευση του εσωτερικού γινομένου δύο διανυσμάτων, που ορίστηκε στον Ορισμό 3.1.3. Έστω $\mathbf{A}, \mathbf{B} \in \mathbb{C}^{I_1 \times I_2}$ τανυστές δεύτερης τάξης, τότε ισχύει η ακόλουθη ιδιότητα:

$$\langle \mathbf{A}, \mathbf{B} \rangle = \sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} a_{i_1 i_2} b^*_{i_1 i_2} = \mathrm{tr}(\mathbf{B}^H \mathbf{A}). \tag{3.27}$$

όπου $\mathrm{tr}(\cdot)$ συμβολίζει το ίχνος ενός πίνακα. Παρατηρούμε ότι το βαθμωτό γινόμενο δύο τανυστών ισούται με το εσωτερικό τους γινόμενο, ως προς όλους τους (κοινούς) δείκτες. Με βάση τον ορισμό του βαθμωτού γινομένου, μπορεί να οριστεί και η γενίκευση της νόρμας Frobenius για τανυστές.

Ορισμός 3.3.9. (**Νόρμα Frobenius Τανυστών**) Θεωρούμε τανυστή $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \ldots \times I_N}$. Η νόρμα Frobenius του $\mathcal{A}$ ορίζεται ως:

$$\|\mathcal{A}\|_F = \sqrt{\langle \mathcal{A}, \mathcal{A} \rangle}. \tag{3.28}$$

□

Η νόρμα Frobenius μπορεί να χρησιμοποιηθεί σαν δείκτης για τη μέτρηση του μεγέθους ενός τανυστή. Το τετράγωνο της νόρμας εκφράζει την ενέργεια του τανυστή.

3.3.7 Ορθογωνιότητα Τανυστών

Ο παρακάτω ορισμός των *αμοιβαίως ορθογωνίων* (mutually orthogonal) τανυστών αποτελεί γενίκευση του αντίστοιχου ορισμού για διανύσματα.

Ορισμός 3.3.10. (**Αμοιβαίως Ορθογώνιοι Τανυστές**) Θεωρούμε τανυστές $\mathcal{A}, \mathcal{B} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_N}$. Οι $\mathcal{A}, \mathcal{B}$ είναι αμοιβαία ορθογώνιοι ($\mathcal{A} \perp \mathcal{B}$) αν ισχύει:

$$\langle \mathcal{A}, \mathcal{B} \rangle = 0. \tag{3.29}$$

□

Στην συνέχεια, θα περιγραφούν έννοιες της ορθογωνιότητας για μία συγκεκριμένη ομάδα τανυστών, τους *αποσυντεθειμένους* (decomposed) τανυστές.

Ορισμός 3.3.11. (**Αποσυντεθειμένος Τανυστής**) Θεωρούμε τανυστή $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_N}$. Ο $\mathcal{A}$ ονομάζεται αποσυντεθετέος, αν μπορεί να αναλυθεί ως:

$$\mathcal{A} = a^{(1)} \otimes a^{(2)} \otimes \cdots \otimes a^{(N)}. \tag{3.30}$$

όπου $a^{(i)} \mathbb{C}^{I_i}$ διανύσματα $i = 1, 2, \ldots, N$ και $\otimes$ συμβολίζει το γινόμενο Kronecker. □

Παράδειγμα 3.3.2. *Θεωρούμε τανυστή* $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$, *με* $\mathcal{A} = (\mathbf{A}_{1ij}|\mathbf{A}_{2ij}|\mathbf{A}_{3ij})$, *όπου:*

$$A_{1ij} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad A_{2ij} = \begin{pmatrix} -1 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{pmatrix} \quad A_{3ij} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Ο $\mathcal{A}$ είναι αποσυνθετέος, καθώς μπορεί να γραφεί ως το εξωτερικό γινόμενο τριών διανυσμάτων: $a^{(1)} = (1, 0, 1),\ a^{(2)} = (0, 1, 0),\ a^{(3)} = (-1, 0, 1)$ ως $\mathcal{A} = a^{(1)} \otimes a^{(2)} \otimes a^{(3)}$. □

Ορισμός 3.3.12. (**Πλήρως Ορθογώνιοι Τανυστές**) Θεωρούμε τανυστές $\mathcal{A}, \mathcal{B} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_N}$ οι οποίοι μπορούν να αποσυντεθούν σε γινόμενα Kronecker διανυσμάτων (αποσυνθετέοι τανυστές). Οι $\mathcal{A}, \mathcal{B}$ είναι πλήρως ορθογώνιοι (completely orthogonal) αν:

$$a^{(i)} \perp b^{(i)} \Leftrightarrow a^{(i)} \cdot b^{(i)} = 0, \quad i = 1, \ldots, N. \tag{3.31}$$

οπότε σημειώνουμε οτι $\mathcal{A} \perp_c \mathcal{B}$. □

Θεωρούμε τον τανυστή $\mathcal{A}$ του Παραδείγματος 3.3.2, καθώς και τον αποσυνθετέο τανυστή $\mathcal{B} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$. Ο $\mathcal{B}$ αποσυντίθεται των διανύσματων $b^{(1)} = (0, 1, 0),\ a^{(2)} = (1, 0, 1),\ a^{(3)} = (0, 1, 0)$. Παρατηρούμε ότι $a^{(1)} \perp b^{(1)}, a^{(2)} \perp b^{(2)}, a^{(3)} \perp b^{(3)}$. Άρα προκύπτει το συμπέρασμα ότι οι $\mathcal{A}$ και $\mathcal{B}$ είναι πλήρως ορθογώνιοι. □

Ορισμός 3.3.13. (**Ισχυρά Ορθογώνιοι Τανυστές**) Θεωρούμε τανυστές $\mathcal{A}, \mathcal{B} \in \mathbb{C}^{I_1 \times \times I_2 \times \cdots \times I_N}$ οι οποίοι μπορούν να αποσυντεθούν σε γινόμενα Kronecker διανυσμάτων. Οι $\mathcal{A}, \mathcal{B}$ είναι ισχυρώς ορθογώνιοι (strongly orthogonal), αν:

$$a^{(i)} \perp b^{(i)} \text{ ή } a^{(i)} = \pm b^{(i)}, \quad i = 1, \ldots, N. \tag{3.32}$$

οπότε σημειώνουμε οτι: $\mathcal{A} \perp_s \mathcal{B}$. □

Θεωρούμε τον τανυστή $\mathcal{A}$ του Παραδείγματος 3.3.2, καθώς και τον αποσυνθετέο τανυστή $\mathcal{B} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$. Ο $\mathcal{B}$ έχει σαν βάση τα διανύσματα $b^{(1)} = (0, 1, 0),\ b^{(2)} = (0, -1, 0),\ b^{(3)} = (-1, 0, 1)$. Παρατηρούμε ότι $a^{(1)} \perp b^{(1)}, a^{(2)} = -b^{(2)}, a^{(3)} = b^{(3)}$. Άρα προκύπτει το συμπέρασμα ότι οι $\mathcal{A}$ και $\mathcal{B}$ είναι ισχυρώς ορθογώνιοι. □

Με βάση τους Ορισμούς 3.3.10-3.3.13, προκύπτει η παρακάτω ιδιότητα για δύο αποσυνθετέους τανυστές $\mathcal{A}, \mathcal{B} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_N}$:

$$\mathcal{A} \perp_c \mathcal{B} \Rightarrow \mathcal{A} \perp_s \mathcal{B} \Rightarrow \mathcal{A} \perp \mathcal{B}. \tag{3.33}$$

Ο παρακάτω ορισμός περιγράφει την έννοια του υπο-τανυστή, που χρησιμοποιείται στον ορισμό ενός ορθογώνιου τανυστή. Σε αντιστοιχία με τον ορισμό ενός ορθογώνιου πίνακα (του οποίου όλα τα διανύσματα που τον συγκροτούν είναι ορθογώνια ανά δύο), η συνθήκη που χρησιμοποιείται είναι οι υπο-τανυστές (κατάντιστοιχεία με τους υποπίνακες) να είναι ορθογώνιοι ανά δύο.

Ορισμός 3.3.14. (**Υπο-τανυστής**) Ο υπο-τανυστής (subtensor) $\mathcal{A}_{i_n=\lambda} \in \mathbb{C}^{I_1\times\cdots\times I_{n-1}\times \times I_{n+1}\times\cdots\times I_N}$ ενός τανυστή $\mathcal{A} \in \mathbb{C}^{I_1\times I_2\times\cdots\times I_N}$ δημιουργείται τον δείκτη i_n ίσο με λ. □

Ορισμός 3.3.15. (**Όλο-Ορθογώνιος Τανυστής**) Θεωρούμε τανυστή $\mathcal{A} \in \mathbb{C}^{I_1\times I_2\times\cdots\times I_N}$. Ο $\mathcal{A}$ είναι όλο-ορθογώνιος (all-orthogonal), αν για κάθε λ, μ, με $\lambda \neq \mu$, ισχύει ότι:

$$\langle \mathcal{A}_{i_n=\lambda}, \mathcal{A}_{i_n=\mu} \rangle = 0. \tag{3.34}$$

□

Στο [66] δίνονται οι ορισμοί των ορθογωνίων τανυστών, μαζί με βασικές ιδιότητές τους. Επίσης, προτείνονται αλγόριθμοι για την εύρεση *ορθογώνιας αποσύνθεσης τανυστών*, δηλαδή την αναπαράσταση ενός τανυστή ως άθροισμα ορθογωνίων τανυστών (Βλ. Ενότητα 3.3.9).

3.3.8 Ανάπτυγμα Τανυστή

Πολλά προβλήματα στην ανάλυση τανυστών είναι δυνατόν να επιλυθούν τοποθετώντας τα στοιχεία των τανυστών σε πίνακες αντίστοιχου μεγέθους. Βασικός λόγος είναι ότι η αναπαράσταση ενός τανυστή σε έναν πίνακα είναι πιο κατανοητή από προγραμματιστική αλλά και διαισθητική άποψη. Επίσης, πολλές γλώσσες προγραμματισμού και πακέτα ανάλυσης δεδομένων δεν υποστηρίζουν πολυδιάστατες δομές. Πολλές από τις ιδιότητες της πολυγραμμικής άλγεβρας έχουν αντιστοιχίες στον δισδιάστατο χώρο της γραμμικής άλγεβρας μετατρέποντας την επίλυση ενός προβλήματος στον τανυστικό χώρο σε επίλυση πολλών υπο-προβλημάτων στον χώρο των πινάκων.

Ορισμός 3.3.16. (**Ανάπτυγμα Τανυστή**) Θεωρούμε τανυστή $\mathcal{A} \in \mathbb{C}^{I_1\times I_2\times\cdots\times I_N}$. Το ανάπτυγμα του $\mathcal{A}$ ως προς τον n-οστό δείκτη είναι ο πίνακας $\mathbf{A}_{(n)} \in \mathbb{C}^{I_n\times(I_{n+1}I_{n+2}\ldots I_N I_1 I_2\ldots I_{n-1})}$. Το στοιχείο $a_{i_1 i_2\ldots i_N}$ του $\mathcal{A}$ βρίσκεται στην i_n σειρά του $\mathbf{A}_{(n)}$, και στην στήλη:

$$\begin{aligned}&(i_{n+1}-1)I_{n+2}I_{n+3}\ldots I_N I_1 I_2\ldots I_{n-1} + (i_{n+2}-1)I_{n+3}I_{n+4}\ldots I_N I_1 I_2\ldots I_{n-1} + \ldots \\ &\quad +(i_N-1)I_1 I_2\ldots I_{n-1} + (i_1-1)I_2 I_3\ldots I_{n-1} + (i_2-1)I_3 I_4\ldots I_{n-1} + \ldots + i_{n-1}.\end{aligned}$$

□

Θεωρούμε τον τανυστή $\mathcal{A} \in \mathbb{C}^{I_1\times I_2\times I_3}$. Ο $\mathcal{A}$ έχει τιμές $a_{ij1} = 1, a_{ij2} = 2, a_{ij3} = 3$ $(i, j = 1, 2, 3)$.

Το ανάπτυγμα $\mathbf{A}_{(1)}$ του $\mathcal{A}$ είναι πίνακας διαστάσεων $I_1 \times I_2 I_3 = 3 \times 3 \cdot 3 = 3 \times 9$:

$$\mathbf{A}_{(1)} = \left(\begin{array}{ccc|ccc|ccc} 1 & 2 & 3 & 1 & 2 & 3 & 1 & 2 & 3 \\ 1 & 2 & 3 & 1 & 2 & 3 & 1 & 2 & 3 \\ 1 & 2 & 3 & 1 & 2 & 3 & 1 & 2 & 3 \end{array}\right)$$

Το ανάπτυγμα $\mathbf{A}_{(2)}$ του $\mathcal{A}$ είναι πίνακας διαστάσεων $I_2 \times I_3 I_1 = 3 \times 3 \cdot 3 = 3 \times 9$:

$$\mathbf{A}_{(2)} = \left(\begin{array}{ccc|ccc|ccc} 1 & 1 & 1 & 2 & 2 & 2 & 3 & 3 & 3 \\ 1 & 1 & 1 & 2 & 2 & 2 & 3 & 3 & 3 \\ 1 & 1 & 1 & 2 & 2 & 2 & 3 & 3 & 3 \end{array}\right)$$

Το ανάπτυγμα $\mathbf{A}_{(3)}$ του $\mathcal{A}$ είναι πίνακας διαστάσεων $I_3 \times I_1 I_2 = 3 \times 3 \cdot 3 = 3 \times 9$:

$$\mathbf{A}_{(3)} = \left(\begin{array}{ccc|ccc|ccc} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 \\ 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 \end{array}\right)$$

□

Μία πολύ χρήσιμη εφαρμογή του αναπτύγματος τανυστή συναντάται στην περίπτωση *αποσύνθεσης τανυστή* (tensor decomposition). Στην αποσύνθεση, ένας τανυστής $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_N}$ είναι δυνατόν να αποσυντεθεί στο γινόμενο ενός τανυστή $\mathcal{B} \in \mathbb{C}^{J_1 \times J_2 \times \cdots \times J_N}$ επί N πίνακες:

$$\mathcal{A} = \mathcal{B} \times_1 \mathbf{C}^{(1)} \times_2 \mathbf{C}^{(2)} \cdots \times_N \mathbf{C}^{(N)} \tag{3.35}$$

όπου οι πίνακες $\mathbf{C}^{(i)} \in \mathbb{C}^{J_i \times I_i}$. Αν εκφραστεί η σχέση (3.35) σε ανάπτυγμα τανυστή ως προς τον n-οστό δείκτη:

$$\mathbf{A}_{(n)} = \mathbf{C}^{(n)} \cdot \mathbf{B}_{(n)} \cdot [\mathbf{C}^{(1)} \otimes \mathbf{C}^{(2)} \otimes \cdots \otimes \mathbf{C}^{(n-1)} \otimes \mathbf{C}^{(n+1)} \otimes \cdots \otimes \mathbf{C}^{(N)}]^T. \tag{3.36}$$

όπου $\otimes$ συμβολίζει το γινόμενο Kronecker.

3.3.9 Βαθμός Τανυστή

Στο πεδίο της γραμμικής άλγεβρας, ο γραμμοβαθμός ή απλώς βαθμός (rank) ενός πίνακα ισούται με τον αριθμό των γραμμικώς ανεξάρτητων γραμμών ή στηλών του πίνακα. Ένας εναλλακτικός ορισμός του βαθμού είναι ο αριθμός των μη μηδενικών οριακών τιμών (ιδιαζουσών τιμών - singular values) του πίνακα. Στο πεδίο της πολυγραμμικής άλγεβρας, η έννοια του βαθμού γενικεύεται, ανάλογα με τη μορφή του τανυστή. Στην περίπτωση μικτού τανυστή, λόγω των περιορισμών που εφαρμόζονται στην κατασκευή του, ο βαθμός ισούται με τον αριθμό των συμμεταβλητών και αντιμεταβλητών δεικτών. Στην περίπτωση τανυστή N-οστής τάξης (component-free approach), ορίζονται πολλές διαφορετικές έννοιες για το βαθμό τανυστή.

Ορισμός 3.3.17. (**Τανυστής 1ου Βαθμού**) Θεωρούμε τανυστή $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_N}$. Αν ο $\mathcal{A}$ μπορεί να γραφεί ως εξωτερικό γινόμενο N διανυσμάτων (είναι αποσυνθετέος), τότε ο βαθμός του ισούται με 1 (Rank-1 tensor). □

Ο παραπάνω ορισμός βρίσκεται σε απόλυτη αντιστοιχία με τον ορισμό ενός πίνακα με βαθμό 1 στην γραμμική άλγεβρα, δηλαδή τον πίνακα που δημιουργείται από το εξωτερικό γινόμενο δύο διανυσμάτων.

Ορισμός 3.3.18. (**Βαθμός Τανυστή**) Θεωρούμε τανυστή $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_N}$. Έστω ότι ο $\mathcal{A}$ ικανοποιεί από την σχέση:

$$\mathcal{A} = \sum_{i=1}^{r} \lambda_i \cdot \mathcal{U}^{(i)}. \tag{3.37}$$

όπου $\lambda_i > 0, i = 1, \ldots, r$ σταθερές και $\mathcal{U}^{(i)} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_N}$ τανυστές με βαθμό 1. Ο βαθμός του $\mathcal{A}$, ο οποίος συμβολίζεται με $\texttt{rank}(\mathcal{A})$ είναι το ελάχιστο r έτσι ώστε ο $\mathcal{A}$ μπορεί να εκφραστεί σύμφωνα με την (3.37). □

Ορισμός 3.3.19. (**Ορθογώνιος Βαθμός Τανυστή**) Θεωρούμε τανυστή $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_N}$. Έστω $\mathcal{A}$ που ικανοποιεί τη σχέση (3.37). Ο ορθογώνιος βαθμός (orthogonal rank) του $\mathcal{A}$, ο οποίος συμβολίζεται με $\texttt{rank}_{\perp}(\mathcal{A})$, είναι το ελάχιστο r έτσι ώστε ο $\mathcal{A}$ να ικανοποιεί την (3.37) και να ισχύει $\mathcal{U}^{(i)} \perp \mathcal{U}^{(j)},\ \forall i \neq j$. Η παραπάνω αποσύνθεση ονομάζεται ορθογώνια αποσύνθεση βαθμού. □

Ορισμός 3.3.20. (**Ισχυρά Ορθογώνιος Βαθμός Τανυστή**) Θεωρούμε τανυστή $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_N}$. Έστω $\mathcal{A}$ που ικανοποιεί την σχέση (3.37). Ο ισχυρά ορθογώνιος βαθμός (strong orthogonal rank) του $\mathcal{A}$, ο οποίος συμβολίζεται με $\texttt{rank}_{\perp_s}(\mathcal{A})$, είναι το ελάχιστο r έτσι ώστε ο $\mathcal{A}$ να ικανοποιεί την (3.37) και να ισχύει $\mathcal{U}^{(i)} \perp_s \mathcal{U}^{(j)},\ \forall i \neq j$. □

Αξίζει να σημειωθεί ότι στην περίπτωση τανυστών δεύτερης τάξης, η αποσύνθεση βαθμού, η ορθογώνια αποσύνθεση βαθμού και η ισχυρά ορθογώνια αποσύνθεση βαθμού ταυτίζονται με την αποσύνθεση οριακών τιμών του πίνακα [66].

Ορισμός 3.3.21. (**Διανύσματα n-οστού Δείκτη**) Θεωρούμε τανυστή $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_N}$. Τα διανύσματα n-οστού δείκτη του $\mathcal{A}$ έχουν διάσταση I_n και προκύπτουν έχοντας ελεύθερο τον δείκτη i_n $(1 \leqslant i_n \leqslant I_n)$ και κρατώντας τους υπόλοιπους δείκτες σταθερούς. □

Ορισμός 3.3.22. (n-**οστός Βαθμός Τανυστή**) Θεωρούμε τανυστή $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_N}$. Ο n-οστός βαθμός του $\mathcal{A}$ συμβολίζεται με $\texttt{rank}_n(\mathcal{A})$ και είναι η διάσταση του διανυσματικού χώρου που δημιουργείται από τα διανύσματα n-οστού δείκτη του $\mathcal{A}$. □

Ο ορισμός του διανύσματος n-οστού δείκτη αποτελεί γενίκευση των διανυσμάτων γραμμής και διανυσμάτων στήλης ενός πίνακα. Αντίστοιχα, ο n-οστός βαθμός τανυστή αποτελεί γενίκευση του γραμμοβαθμού. Στην περίπτωση του τανυστή όμως, οι n-οστοί βαθμοί δεν ταυτίζονται μεταξύ τους απαραίτητα. Και στην περίπτωση που ταυτίζονται, υπάρχει περίπτωση να διαφέρουν

από τον βαθμό του τανυστή. Η σχέση που ισχύει μεταξύ του βαθμού τανυστή και του n-οστού βαθμού τανυστή είναι: $\texttt{rank}_n(\mathcal{A}) \leqslant \texttt{rank}(\mathcal{A})$.

Οι n-οστοί βαθμοί ενός τανυστή μπορούν να υπολογιστούν εύκολα χρησιμοποιώντας το αντίστοιχο ανάπτυγμα του τανυστή:

$$\texttt{rank}_n(\mathcal{A}) = \texttt{rank}(\mathbf{A}_{(n)}). \tag{3.38}$$

Η εύρεση του βαθμού ενός τανυστή είναι πολύ πιο σύνθετο πρόβλημα σε σχέση με την εύρεση της n-οστού βαθμού, γιατί θεωρεί τον τανυστή σαν μία n-διάστατη ποσότητα. Ο καθορισμός του βαθμού ενός $I_1 \times I_2 \times \cdots \times I_N$ τανυστή είναι ακόμα ανοικτό πρόβλημα στην βιβλιογραφία. Έχουν προταθεί τεχνικές για την εύρεση βαθμού τανυστών συγκεκριμένης τάξης και διαστάσεων (πχ. για τανυστές $2 \times 2 \times \cdots \times 2$). Επίσης, αποδεικνύεται ότι το πρόβλημα εύρεσης του βαθμού ενός τανυστή 3ης τάξης είναι εκθετικής πολυπλοκότητας. Τέλος, αποδεικνύεται ότι η τιμή του βαθμού του τανυστή εξαρτάται από το πεδίο στο οποίο ορίζεται ο τανυστής. Κοινώς, ο βαθμός ενός πραγματικού τανυστή που ορίζεται στο πεδίο των πραγματικών αριθμών δεν έχει ίδια τιμή με τον βαθμό του ίδιου τανυστή αν οριστεί στο μιγαδικό πεδίο. Με την παραπάνω παρατήρηση φαίνεται η διαφορά στην εύρεση του βαθμού ενός τανυστή σε σχέση με την εύρεση του βαθμού ενός πίνακα, όπου έχουν προταθεί πολλές τεχνικές για την εύρεσή της όπως η ανάλυση οριακών τιμών, η αποσύνθεση QR, η απαλοιφή Gauss κτλ.

Θεωρούμε τον τανυστή $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$ του Παραδείγματος 3.3.2. Εφόσον $\mathcal{A} = a^{(1)} \otimes a^{(2)} \otimes a^{(3)}$, ο βαθμός του $\mathcal{A}$ δεν μπορεί να είναι μεγαλύτερος από 1, σύμφωνα με την (3.37). Επειδή προφανώς ο βαθμός ενός τανυστή είναι θετικός ακέραιος, ο βαθμός του $\mathcal{A}$ ισούται με 1. □

Θεωρούμε τον τανυστή $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$. Ο $\mathcal{A}$ έχει τιμές $a_{11j} = a_{22j} = 1, a_{12j} = a_{21j} = 2$ $(j = 1, 2)$.

Το ανάπτυγμα $\mathbf{A}_{(1)}$ του $\mathcal{A}$ είναι ο 2×2 πίνακας:

$$\mathbf{A}_{(1)} = \left(\begin{array}{cc|cc} 1 & 1 & 2 & 2 \\ 2 & 2 & 1 & 1 \end{array}\right)$$

Ο 1-οστός βαθμός του $\mathcal{A}$ ισούται με τον βαθμό του $\mathbf{A}_{(1)}$: $\texttt{rank}_1(\mathcal{A}) = 2$. Το ανάπτυγμα $\mathbf{A}_{(2)}$ του $\mathcal{A}$ είναι ο 2×4 πίνακας:

$$\mathbf{A}_{(2)} = \left(\begin{array}{cc|cc} 1 & 2 & 1 & 2 \\ 2 & 1 & 2 & 1 \end{array}\right)$$

Ο 2-οστός βαθμός του $\mathcal{A}$ ισούται με τον βαθμό του $\mathbf{A}_{(2)}$: $\texttt{rank}_2(\mathcal{A}) = 2$. Το ανάπτυγμα $\mathbf{A}_{(3)}$ του $\mathcal{A}$ είναι ο 2×2 πίνακας:

$$\mathbf{A}_{(3)} = \left(\begin{array}{cc|cc} 1 & 2 & 2 & 1 \\ 1 & 2 & 2 & 1 \end{array}\right)$$

Ο 3-οστός βαθμός του $\mathcal{A}$ ισούται με τον βαθμό του $\mathbf{A}_{(3)}$: $\texttt{rank}_3(\mathcal{A}) = 1$. Θα υπολογίσουμε τώρα τον βαθμό του $\mathcal{A}$: ισχύει ότι $\mathcal{A} = a^{(1)} \otimes a^{(2)} \otimes (a^{(1)} + a^{(3)}) + a^{(3)} \otimes a^{(2)} \otimes (a^{(1)} + a^{(3)})$, όπου $a^{(1)} = (0\ 1)$, $a^{(2)} = (2\ 1)$, και $a^{(3)} = (1\ 0)$. Άρα, ο βαθμός του $\mathcal{A}$ δεν είναι μεγαλύτερος από 2. Εφ'οσον $\texttt{rank}_1(\mathcal{A}) = \texttt{rank}_2(\mathcal{A}) = 2$, ο βαθμός του $\mathcal{A}$ ισούται με 2. □

3.3.10 Υπερ-συμμετρικοί Τανυστές

Στο πεδίο της γραμμικής άλγεβρας, η έννοια της συμμετρίας συναντάται σε δύο εκδοχές: την πραγματική και την ερμιτιανή συμμετρία. Στο πεδίο της πολυγραμμικής άλγεβρας, η πραγματική και ερμιτιανή συμμετρία γενικεύονται στην έννοια της υπερ-συμμετρίας (super-symmetry). Προτείνεται επίσης και η έννοια της συμμετρίας ανά δύο (pairwise symmetry), που αποτελεί μία λιγότερο περιοριστική γενίκευση της συμμετρίας.

Θεωρώντας έναν τανυστή $\mathcal{A} \in \mathbb{R}^{I \times I \times \cdots \times I}$, η έννοια της πραγματικής συμμετρίας μπορεί να γενικευτεί θεωρώντας ότι τα στοιχεία του τανυστή $a_{i_1 i_2 \ldots i_N}$ έχουν ίδια τιμή με στοιχεία που προκύπτουν από οποιαδήποτε αντιμετάθεση των δεικτών $i_1, i_2, \ldots, i_N$. Στην περίπτωση μιγαδικού τανυστή, η γενίκευση δεν είναι τετριμμένη, και σχετίζεται με την αναπαράσταση ομογενών πολυώνυμων στον τανυστικό χώρο.

Ορισμός 3.3.23. (**Σχετιζόμενο Πολυώνυμο Τανυστή**) Θεωρούμε τανυστή $\mathcal{A} \in \mathbb{R}^{I \times I \times \cdots \times I}$ και διάνυσμα $\mathbf{x} \in \mathbb{R}^I$. Το σχετιζόμενο πολυώνυμο (associated polynomial) $\mathcal{A}(\mathbf{x})$ του $\mathcal{A}$ ορίζεται ως:

$$\mathcal{A}(\mathbf{x}) = \mathcal{A} \times_1 \mathbf{x} \times_2 \mathbf{x} \ldots \times_N \mathbf{x} = \sum_{i_1=1}^{I} \sum_{i_2=1}^{I} \ldots \sum_{i_N=1}^{I} a_{i_1 i_2 \ldots i_N} x_{i_1} x_{i_2} \ldots x_{i_N}. \tag{3.39}$$

Στην περίπτωση μιγαδικού τανυστή $\mathcal{A} \in \mathbb{C}^{I \times I \times \cdots \times I}$, το σχετιζόμενο πολυώνυμο ορίζεται ως:

$$\mathcal{A}(\mathbf{x}) = \mathcal{A} \times_1 \tilde{\mathbf{x}}^{(1)} \times_2 \tilde{\mathbf{x}}^{(2)} \ldots \times_N \tilde{\mathbf{x}}^{(N)} = \sum_{i_1=1}^{I} \sum_{i_2=1}^{I} \ldots \sum_{i_N=1}^{I} a_{i_1 i_2 \ldots i_N} \tilde{x}^{(1)}_{i_1} \tilde{x}^{(2)}_{i_2} \ldots \tilde{x}^{(N)}_{i_N}. \tag{3.40}$$

όπου τα διανύσματα $\tilde{\mathbf{x}}^{(n)}$ ισούνται είτε με το $\mathbf{x}^{(n)}$ είτε με το $\mathbf{x}^{*(n)}$ $(n = 1, 2, \ldots, N)$. □

Ορισμός 3.3.24. (**Υπερ-συμμετρικός Τανυστής**) Θεωρούμε τανυστή $\mathcal{A} \in \mathbb{C}^{I \times I \times \cdots \times I}$ και το σχετιζόμενο πολυώνυμο $\mathcal{A}(\mathbf{x})$. Ο $\mathcal{A}$ είναι υπερ-συμμετρικός αν: 1) Για κάθε αντιμετάθεση P έτσι ώστε $\tilde{x}^{(1)}_{\mathsf{P}(i_1)} \tilde{x}^{(2)}_{\mathsf{P}(i_2)} \ldots \tilde{x}^{(N)}_{\mathsf{P}(i_N)} = \tilde{x}^{(1)}_{i_1} \tilde{x}^{(2)}_{i_2} \ldots \tilde{x}^{(N)}_{i_N}$, ισχύει ότι $a_{\mathsf{P}(i_1 i_2 \ldots i_N)} = a_{i_1 i_2 \ldots i_N}$. 2) Για κάθε αντιμετάθεση P έτσι ώστε $\tilde{x}^{(1)}_{\mathsf{P}(i_1)} \tilde{x}^{(2)}_{\mathsf{P}(i_2)} \ldots \tilde{x}^{(N)}_{\mathsf{P}(i_N)} = (\tilde{x}^{(1)}_{i_1} \tilde{x}^{(2)}_{i_2} \ldots \tilde{x}^{(N)}_{i_N})^*$ ισχύει ότι $a_{\mathsf{P}(i_1 i_2 \ldots i_N)} = a^*_{i_1 i_2 \ldots i_N}$. □

Όσον αφορά μικτούς τανυστές, ορίζεται η οριζόντια συμμετρία (horizontal symmetry) ως ανταλλαγή δύο συμμεταβλητών ή αντιμεταβλητών δεικτών του τανυστή. Αντίστοιχα, η κάθετη

συμμετρία (vertical symmetry) ορίζεται ως ανταλλαγή ανάμεσα σε έναν συμμεταβλητό και έναν αντιμεταβλητό δείκτη. Η υπερ-συμμετρία σε μικτούς τανυστές ορίζεται ως συνδυασμός οριζόντιας και κάθετης συμμετρίας. Σε έναν υπερ-συμμετρικό τανυστή με p συμμεταβλητούς και p αντιμεταβλητούς δείκτες, πρέπει να ισχύουν $2 \cdot (p!)^2$ οριζόντιες και κάθετες συμμετρίες.

Αξίζει να σημειωθεί ότι το σχετικό πολυώνυμο ενός υπερ-συμμετρικού τανυστή παίρνει μόνο πραγματικές τιμές. Επίσης, ο ορισμός της υπερ-συμμετρίας είναι ανεξάρτητος από την βάση στην οποία ορίζεται το διάνυσμα $\mathbf{x}$. Οι υπερ-συμμετρικοί τανυστές έχουν μεγάλη χρήση στο πεδίο της στατιστικής υψηλής τάξης (higher-order statistics), που θα περιγραφεί συνοπτικά σε προσεχή ενότητα. Ακολουθεί ένας πιο χαλαρός ορισμός της συμμετρίας τανυστών, η συμμετρία τανυστή ανά δύο.

Ορισμός 3.3.25. (**Συμμετρικός Τανυστής Ανά Δύο**) Θεωρούμε τανυστή $\mathcal{A} \in \mathbb{C}^{I \times I \times \cdots \times I}$. Ο $\mathcal{A}$ είναι συμμετρικός ανά δύο αν, για κάθε ζεύγος δεικτών (i_{n_1}, i_{n_2}), υπάρχει μία αντιμετάθεση $\mathrm{P}_{n_1 n_2}$ έτσι ώστε $a_{\mathrm{P}_{n_1 n_2}(i_1 i_2 \ldots i_N)} = a_{i_1 i_2 \ldots i_N}$ ή $a_{\mathrm{P}_{n_1 n_2}(i_1 i_2 \ldots i_N)} = a^*_{i_1 i_2 \ldots i_N}$. □

Θεωρούμε τον τανυστή $\mathcal{A} \in \mathbb{R}^{2 \times 2 \times 2}$. Ο $\mathcal{A}$ παίρνει τις τιμές:

$$a_{ij1} = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix} \quad a_{ij2} = \begin{pmatrix} 2 & 3 \\ 3 & 4 \end{pmatrix}$$

Παρατηρούμε ότι ισχύει η ακόλουθη σχέση: $a_{i_1 i_2 i_3} = a_{i_2 i_3 i_1}$. Επίσης παρατηρείται ότι η παραπάνω σχέση είναι δυνατόν να ισχύσει αν γίνουν δύο αντιμεταθέσεις δεικτών στον $a_{i_1 i_2 i_3}$: (i_1, i_2) και (i_1, i_3). Άρα, ο $\mathcal{A}$ είναι συμμετρικός ανά δύο. □

3.3.11 Τανυστές ως Γραμμικοί Μετασχηματισμοί

Οι τανυστές υψηλής τάξης μπορούν να χρησιμοποιηθούν για μετασχηματισμούς ανάμεσα σε διανυσματικούς χώρους (και τανυστικούς χώρους), για παράδειγμα ανάμεσα σε διανύσματα, ή ανάμεσα σε έναν πίνακα και έναν τανυστή 3ης τάξης. Προφανώς, η συζήτηση αποτελεί μία γενίκευση των γραμμικών μετασχηματισμών με πίνακες, οι οποίοι μπορούν να χρησιμοποιηθούν για να συνδέσουν διανυσματικούς χώρους.

Ορισμός 3.3.26. (**Γραμμικός Μετασχηματισμός Τανυστών**) Θεωρούμε τους τανυστές $\mathcal{A} \in \mathbb{C}^{J_1 \times J_2 \times \cdots \times J_{N_1}}$ και $\mathcal{B} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_{N_2}}$. Ο γραμμικός μετασχηματισμός του $\mathcal{A}$ στον $\mathcal{B}$ πραγματοποιείται χρησιμοποιώντας τον τανυστή $\mathcal{C} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_{N_2} \times J_1 \times J_2 \times \cdots \times J_{N_1}}$:

$$\mathcal{B} = \langle \mathcal{C}, \mathcal{A} \rangle_{N_2+1,\ldots,N_2+N_1;1,\ldots,N_1}. \tag{3.41}$$

Η σχέση (3.41) μπορεί να γραφεί εναλλακτικά:

$$b_{i_1 i_2 \ldots i_{N_2}} = \sum_{j_1=1}^{J_1} \sum_{j_2=1}^{J_2} \cdots \sum_{j_{N_1}=1}^{J_{N_1}} c_{i_1 i_2 \ldots i_{N_2} j_1 j_2 \ldots j_{N_1}} a_{j_1 j_2 \ldots j_{N_1}}. \tag{3.42}$$

□

Στον παραπάνω ορισμό, ο τανυστής $\mathcal{C}$ μπορεί να χρησιμοποιηθεί για να αναπαραστήσει πολλές διαφορετικές συνδέσεις, ανάλογα με τους δείκτες που χρησιμοποιούνται στα αθροίσματα της (3.42).Σημειωτέον ότι ο παραπάνω μετασχηματισμός μπορεί να υλοποιηθεί χρησιμοποιώντας τα αναπτύγματα των τανυστών. Σ' αυτήν την περίπτωση, το ανάπτυγμα του $\mathcal{B}$ μπορεί να υπολογιστεί με το γινόμενο των πινάκων που εκφράζουν τα αναπτύγματα των $\mathcal{A}$ και $\mathcal{C}$. Να σημειωθεί επίσης ότι στην περίπτωση των μικτών τανυστών, ο μετασχηματισμός γίνεται ανάμεσα σε συμμεταβλητούς και αντιμεταβλητούς δείκτες των $\mathcal{A}$ και $\mathcal{C}$.

Κεφάλαιο 4

Αναγνώριση Μουσικού Είδους

Στο παρόν κεφάλαιο εισάγεται ο όρος *αναγνώριση ή ταξινόμηση μουσικού είδους* (music genre classification) και περιγράφεται πώς αυτή συντελείται από τους ανθρώπους και από τις μηχανές.

Το μουσικό είδος χρησιμοποιείται από τους ανθρώπους ως σύντομη περιγραφή του περιεχομένου ενός μουσικού κομματιού. Η πληροφορία του μουσικού είδους εκτός από τη μουσική δομή, συχνά εμπεριέχει και πληροφορίες που σχετίζονται με την ιστορία και την χρονική περίοδο, καθώς και με το κοινωνικό και πολιτισμικό περιβάλλον κατά το οποίο δημιουργήθηκε το μουσικό κομμάτι.

Αναγνώριση μουσικού είδους είναι η διαδικασία κατά την οποία ένα δεδομένο μουσικό κομμάτι αντιστοιχίζεται σε μια κατηγορία μουσικού είδους π.χ. jazz, rock ή pop κ.α. Διαφορετικά μουσικά κομμάτια που ανήκουν στο ίδιο μουσικό είδος, μοιράζονται την ίδια 'βασική μουσική γλώσσα' [92] ή προέρχονται από το ίδιο πολιτισμικό περιβάλλον ή ακόμη και από την ίδια ιστορική περίοδο.

Η *αυτόματη αναγνώριση μουσικού είδους* έγκειται στην ταξινόμηση των μουσικών κομματιών σε διακριτά μουσικά είδη με τη χρήση ηλεκτρονικού υπολογιστή. Ως ερευνητική περιοχή η αναγνώριση μουσικού είδους είναι πολύ δημοφιλής αν και έχει ιστορία μόλις μιας δεκαετίας. Χαρακτηριστικό αυτής της ερευνητικής περιοχής είναι η διεπιστημονική προσέγγιση του προβλήματος, επειδή η επίλυσή του απαιτεί γνώσεις από διαφορετικά και ετερόκλητα επιστημονικά πεδία όπως η ψηφιακή επεξεργασία σήματος, η μουσική θεωρία, η ψυχοφυσιολογία και ψυχοακουστική, η μηχανική μάθηση και η αναγνώριση προτύπων.

Τα κίνητρα για την ανάπτυξη αλγορίθμων αναγνώρισης μουσικού είδους είναι τόσο εμπορικά όσο και επιστημονικά. Για παράδειγμα η υπηρεσία iTunes της Apple διαχειρίζεται μια βάση δεδομένων που περιλαμβάνει περισσότερα από 2.000.000 αρχεία μουσικής. Η χειροκίνητη οργάνωση και ταξινόμηση σε μουσικά είδη ενός τόσο μεγάλου πλήθους αρχείων είναι εξαιρετικά χρονοβόρα και ακριβή διαδικασία. Επομένως εργαλεία αυτόματης ταξινόμησης είναι αναγκαί-

α. Εκτός από τα εμπορικά κίνητρα, η έρευνα στο πεδίο της αυτόματης ταξινόμησης μουσικού είδους παρακινείται και από τη στενή σχέση του πεδίου με αλλά παρεμφερή ερευνητικά πεδία της περιοχής της Ανάκτησης Μουσικής Πληροφορίας (Music Information Retrieval - MIR). Η ερευνητική περιοχής της Ανάκτησης Μουσικής Πληροφορίας καλύπτει σχεδόν το σύνολο των διαδικασιών που σχετίζονται με την διαχείριση του ψηφιακού μουσικού υλικού καθώς επίσης και προβλήματα στενά συνδεδεμένα με την αυτόματη ταξινόμηση μουσικού είδους. Για παράδειγμα, τα προβλήματα της αναγνώρισης καλλιτέχνη (music artist identification), της αναγνώρισης μουσικών οργάνων, της αναγνώρισης ρυθμού, του ακουστικού αποτυπώματος audio fingerprinting και της αναγνώρισης μουσικού είδους είναι στενά συσχετισμένα υπό την έννοια ότι η επίλυση τους εξαρτάται από την ποιότητα της αναπαράστασης του μουσικού σήματος μέσω χαρακτηριστικών τα οποία είναι συμπαγή (compact) και εμπεριέχουν αρκετή πληροφορία για το μουσικό σήμα. Συνεπώς, τα χαρακτηριστικά που είναι κατάλληλα για την αναγνώριση μουσικού είδους πιθανό να είναι χρήσιμα και για τα προαναφερθέντα προβλήματα και αντιστρόφως.

4.1 Αναγνώριση Μουσικού Είδους από τον Άνθρωπο

Το ανθρώπινο σύστημα ακοής εξελίσσεται με την εξέλιξη του ανθρώπινου είδους. Έτσι ενώ αρχικά η λειτουργία του περιοριζόταν στο να εντοπίζει ηχητικές πηγές στο διάβα των αιώνων εξελίχτηκε τόσο, ώστε να ερμηνεύει τις πολύπλοκες γλωσσικές δομές και να διαδραματίζει σπουδαίο ρόλο στην επικοινωνία του ανθρώπου. Η μουσική έχει μακρά ιστορία περίπου όσο και η ομιλία και χρονολογείται από τη προϊστορική εποχή. Το πρώτο γνωστό μουσικό όργανο χρονολογείται στα 80.000-40.000 π.Χ. και έμοιαζε με αυλό κατασκευασμένο από κέρατο ζώου. Η μουσική σχετίζεται έμμεσα με την ομιλία υπό την έννοια ότι παράγεται από ανθρώπους με χαρακτηριστικά τέτοια, ώστε να γίνεται αντιληπτή από το ανθρώπινο σύστημα ακοής. Επιπρόσθετα, η μουσική συχνά περιέχει τραγούδι το οποίο σχετίζεται άμεσα με την ομιλία και τη διαδικασία παραγωγής και αντίληψης της. Σύμφωνα με τον παραπάνω συσχετισμό μουσικής και ομιλίας, τα αποτελέσματα των ερευνών σχετικά με την παραγωγή, την αντίληψη και την μοντελοποίηση της ομιλίας, μπορούν να χρησιμοποιηθούν και στην κατανόηση αντίστοιχων διαδικασιών που σχετίζονται με τη μουσική.

Ο φυσικός τρόπος παραγωγής μουσικής πραγματοποιείται από τον άνθρωπο χρησιμοποιώντας την φωνή σε συνδυασμό με μουσικά όργανα τα οποία γενικά ανήκουν σε τρεις κατηγορίες (πνευστά, έγχορδα και κρουστά) σύμφωνα με τον τρόπο που αυτά παράγουν τον ήχο. Παρόλα αυτά τα τελευταία χρόνια η παραγωγή μουσικής έχει αναθεωρηθεί, και η σύγχρονη μουσική

περιέχει και ακουστικά στοιχεία τα οποία δεν παράγονται από φυσικά μουσικά όργανα, αλλά από ηλεκτρονικά.

Το βασικά αντιλαμβανόμενα χαρακτηριστικά της μουσικής περιγράφονται από την μουσική θεωρία. Παραδοσιακά τα χαρακτηριστικά αυτά είναι εκείνα του δυτικού μουσικού σύστηματος (δηλαδή της ευρωπαϊκής κλασσικής μουσικής) όπως η μελωδία, η αρμονία, ο ρυθμός και το ηχόχρωμα. Τα προαναφερθέντα χαρακτηριστικά σχετίζονται αρκετά με το σύστημα σημειογραφίας της δυτικής μουσικής [33].

Πολλές φορές, η μουσική περιγράφεται με όρους όπως η υφή και το στυλ, οι οποίοι μπορεί να θεωρηθούν ως συνδυασμός βασικών χαρακτηριστικών. Η παραπάνω παρατήρηση ισχύει διότι η αντίληψη της μουσικής οφείλεται σε γνωστικές διαδικασίες υψηλού επιπέδου. Επιπλέον ο συνδυασμός των βασικών χαρακτηριστικών είναι προσθετικός, γεγονός που λαμβάνεται υπόψη στην παρούσα διατριβή.

Η ακριβής διαδικασία αναγνώρισης του μουσικού είδους από τους ανθρώπους είναι ακόμη άγνωστη. Ενδεχόμενως η διαδικασία αυτή να περιλαμβάνει βιοφυσικές και γνωστικές διαδικασίες, που αξιοποιούν την πληροφορία που περιέχουν τα βασικά χαρακτηριστικά της μουσικής. Όμως, χαρακτηριστικά που δεν ανήκουν στην μουσική αυτή καθαυτή επηρεάζουν την ταξινόμηση. Το πολιτισμικό και ιστορικό υπόβαθρο του ατόμου επηρεάζει σημαντικά πώς ταξινομεί τη μουσική σε είδη. Σημαντικό ρόλο στη διαδικασία αυτή κατέχουν οι δισκογραφικές εταιρίες και τα δισκοπωλεία, αφού αυτά πρώτα ταξινομούν τη μουσική σε μουσικά είδη επηρεάζοντας ή ακόμη και διαμορφώνοντας πολλές φορές την άποψη των ακροατών.

Είναι σαφές ότι η ταξινόμηση της μουσικής σε είδη επηρεαζεται από πολλούς και διαφορετικούς παράγοντες. Παρόλα αυτά δεν είναι σαφής η σημασία κάθε παράγοντα. Ασαφής είναι και η σχέση των εσωτερικών χαρακτηριστικών της μουσικής με τους εξωτερικούς παράγοντες που επηρεάζουν την ταξινόμηση της σε είδη, δηλαδή τα χαρακτηριστικά του ακουστικού σήματος σε σχέση με πολιτισμικούς, ιστορικούς και λοιπούς παράγοντες. Μια απάντηση στην παραπάνω ερώτηση προκύπτει από τα αποτελέσματα σχετικών πειράματων στο [24]. Σ' αυτά τα πειράματα τρία ψάρια - κυπρίνοι - εκπαιδευτήκαν ώστε να ταξινομούν την μουσική που άκουγαν σε Blues ή κλασσική. Το σύστημα ακοής των κυπρίνων έχει εξαιρετικές ομοιότητες με το ανθρώπινο. Κατόπιν εκπαίδευσης, τα ψάρια κατάφεραν να ταξινομούν σωστά τη μουσική στα διακριτά είδη Blues και κλασσική γεγονός που καταδεικνύει ότι η εσωτερική δομή της μουσικής παρέχει αρκετή πληροφορία έτσι ώστε να διακριθούν δυο τόσο διαφορετικά μουσικά είδη. Προφανώς τα ψάρια δεν γνωρίζουν το πολιτισμικό περιβάλλον των δυο διαφορετικών μουσικών ειδών.

Στο [7] εξετάζονται οι δυνατότητες του ανθρώπου να ταξινομεί σωστά τη μουσική σε μουσικά στυλ (υπό-είδη). Πιο συγκεκριμένα τα 4 στυλ που εξετάζονται ανήκουν σε 4 ιστορικές περιόδους της κλασσικής μουσικής, από το baroque μέχρι το νεορομαντισμό. Στα πειράματα

εξετάστηκε η υπόθεση της *ιστορικής απόστασης*. Η υπόθεση διατυπώνεται ως εξής. Η μουσική η οποία δημιουργήθηκε σε κοντινές χρονικές περιόδους θα έχει αρκετές ομοιότητες. Η υπόθεση επιβεβαιώθηκε. Το ακροατήριο που χρησιμοποιήθηκε στα πειράματα περιλάμβανε δυτικούς μουσικούς και μη μουσικούς, όπως και μη δυτικούς μουσικούς και μη μουσικούς. Είναι ενδιαφέρον ότι ακόμη και οι μη δυτικοί που δεν είχαν ποτέ ακούσει κλασσική μουσική κατάφεραν να διακρίνουν το αποτέλεσμα της ιστορικής απόστασης. Μπορούμε να συμπεράνουμε λοιπόν ότι το πολιτισμικό υπόβαθρο δεν επηρεάζει την αναγνώριση του μουσικού είδους στους ανθρώπους. Η ομάδα των δυτικών, μουσικών και μη, είχε καλύτερη επίδοση στην αναγνώριση του μουσικού υπο-είδους από τους μη δυτικούς, πράγμα που σημαίνει ότι απλά και μονό η έκθεση σε δυτικά ακούσματα αυξάνει την ικανότητα διάκρισης των μουσικών στυλ χωρίς το υποκείμενο να έχει συστηματική μουσική παιδεία. Είναι αναμενόμενο ότι οι δυτικοί μουσικοί είχαν τις καλύτερες επιδόσεις.

4.1.1 Το Πρόβλημα Ορισμού του Μουσικού Είδους

Το πρόβλημα της αναγνώρισης μουσικών ειδών δεν είναι τετριμμένο. Ενώ υπάρχει διαδεδομένη χρήση των μουσικών ειδών, αυτά παραμένουν μία κακώς ορισμένη έννοια. Είναι χαρακτηριστικό ότι η σωστή ταξινόμηση μουσικού είδους από ανθρώπους (μη ειδικούς στην ταξινόμηση μουσικών ειδών) φτάνει πολύ χαμηλά ποσοστά: για το πρόβλημα της ταξινόμησης κομματιών σε 10 μουσικά είδη, η ανθρώπινη ακρίβεια είναι μόλις 53% για δείγματα 250 msec, και ανεβαίνει στο 72% για δείγματα 3 sec [101]. Ταυτόχρονα, ενώ πολύ συχνά χρησιμοποιούνται όροι, όπως pop, rock, και jazz, αυτοί είναι ασαφώς ορισμένοι και πολλές φορές ένα κομμάτι αντιστοιχεί σε περισσότερα από ένα μουσικά είδη. Τέλος, τίθεται το ζήτημα πού θα εφαρμοστεί η ταξινόμηση μουσικού είδους: στο κομμάτι, στο άλμπουμ (συλλογή κομματιών) ή στον καλλιτέχνη. Στα περισσότερα ηλεκτρονικά καταστήματα η ταξινόμηση γίνεται με βάση το άλμπουμ. Όμως, γίνεται κατανοητό ότι αφού ένα κομμάτι μπορεί να ανήκει σε περισσότερα από ένα είδη, τότε ένα άλμπουμ, το οποίο μπορεί και να περιέχει ετερογενές υλικό, δεν είναι δυνατόν να ανήκει σε ένα μόνο είδος.

Άρα, για το πρόβλημα της ταξινόμησης μουσικών ειδών, απαιτείται μία ιεραρχική ταξινόμηση των μουσικών ειδών, η οποία να καλύπτει όσο το δυνατόν γίνεται όλα τα είδη, και παράλληλα να μην υπάρχουν πολλές επικαλύψεις και ασάφειες στους ορισμούς των ειδών. Ο Pachet όμως έδειξε ότι η δημιουργία μίας τέτοιας ιεραρχίας δεν είναι απλό ζήτημα [95]. Για παράδειγμα, το site Amazon[1] έχει κατηγοριοποιήσει κομμάτια χρησιμοποιώντας 719 μουσικά είδη. Το site All-

[1] `http://www.amazon.com`

music[2] χρησιμοποιεί 531 είδη και το Mp3 [3] χρησιμοποιεί 430 είδη. Μόνο 70 μουσικά είδη ήταν κοινά και στα τρία sites. Επίσης, παρατηρείται από τον Pachet ότι οι ασάφειες και διαφορετικές ταξινομήσεις ειδών δεν αποτελούν κατ' ανάγκη πρόβλημα για τους χρήστες, αλλά προφανώς δεν προσφέρονται για χρήση σε συστήματα αυτόματης ταξινόμησης [95]. Ένα άλλο πρόβλημα προστίθεται στο πρόβλημα εύρεσης ταξινόμησης: ότι εξαρτάται από την περιοχή στην οποία προτείνεται. Για παράδειγμα, ένα κομμάτι έντεχνης ελληνικής μουσικής σε μη ελληνικές υπηρεσίες θα ταξινομηθεί ως μουσική του κόσμου (world music). Επίσης, υπάρχει ασάφεια με βάση τα κριτήρια από τα οποία προκύπτει μία ιεραρχία: άλλες ιεραρχίες προκύπτουν χαρακτηρίζοντας χρονικές περιόδους (πχ. τραγούδια δεκαετίας του '50), άλλες με βάση τη χώρα, με βάση τη γλώσσα, με βάση το θέμα των κομματιών, ή με βάση τον καλλιτέχνη. Ένα τελευταίο πρόβλημα που προκύπτει, είναι η προσθήκη νέου μουσικού είδους στην ήδη υπάρχουσα ιεραρχία. Τα νέα μουσικά είδη προκύπτουν συνήθως συγχωνεύοντας ήδη υπάρχοντα είδη, ή χωρίζοντας ένα ήδη υπάρχον είδος σε κατηγορίες. Το παραπάνω θέμα αποτελεί πρόβλημα για ένα σύστημα αυτόματης ταξινόμησης μουσικών ειδών, το οποίο θα πρέπει να μεταβάλλει δυναμικά την ιεραρχία του.

4.2 Αυτόματη Αναγνώριση Μουσικού Είδους

Μία πληθώρα αλγορίθμων αυτόματης αναγνώρισης μουσικού είδους έχει προταθεί στην βιβλιογραφία. Η πλειονότητα των αλγορίθμων εξάγουν βασικά χαρακτηριστικά από το μουσικό σήμα και ταξινομούν το κομμάτι σε κάποιο μουσικό είδος χρησιμοποιώντας κάποιο αλγόριθμο μηχανικής μάθησης κατόπιν εκπαίδευσης. Η επίδοση των αλγορίθμων αξιολογείται με πειράματα σε σύνολα δεδομένων κατασκευασμένα για το σκοπό αυτό.

Στην αυτή την ενότητα παρουσιάζουμε αρχικά δημοφιλή σύνολα δεδομένων για την αναγνώριση μουσικού είδους, τα συνήθη χαρακτηριστικά καθώς και τους βασικούς αλγορίθμους μηχανικής μάθησης που χρησιμοποιούνται. Συγκεντρωτικά αποτελέσματα των πιο αξιόλογων προσπαθειών παρουσιάζονται στην ενότητα 4.2.4.

4.2.1 Σύνολα Δεδομένων και Ιεραρχίες

Ο Pachet αποπειράθηκε το 2000 να δημιουργήσει μία εξαντλητική ιεραρχία μουσικών ειδών, προσπαθώντας παράλληλα να μην υπάρχουν αλλεπικαλύψεις σε μουσικά είδη, να δίνεται η δυνατότητα προσθήκης νέου μουσικού είδους, και η ιεραρχία να είναι 'αντικειμενική', δηλαδή να μην επηρεάζεται από χρονικούς και τοπικούς παράγοντες [95]. Η ιεραρχία που δημιουργήθηκε

[2] `http://www.allmusic.com`

[3] `http://www.mp3.com`

είχε σαν πρώτο επίπεδο βασικά μουσικά είδη και σε δεύτερο επίπεδο περιγραφές εξειδίκευσης του μουσικού είδους (οι περιγραφές αναφέρονται σε χώρα, ενορχήστρωση, και καλλιτέχνη). Συνολικά η ιεραρχία που προτάθηκε περιείχε 378 μουσικά είδη. Όμως, τελικά ο Pachet εγκατέλειψε τη δημιουργία της ιεραρχίας [3], λόγω πολλών επικαλύψεων που υπήρχαν ανάμεσα σε είδη και σε ευαισθησία της ταξινόμησης σε νέα μουσικά είδη που προκύπτουν από μείξη ειδών που υπήρχαν στην ιεραρχία. Η νέα ιεραρχία που προτάθηκε από τον Pachet - για το πρόγραμμα αναζήτησης μουσικών ειδών του ερευνητικού έργου Cuidado - είχε 2 επίπεδα, το πρώτο με 18 γενικά μουσικά είδη και το δεύτερο με 250 υπο-είδη [96]. Τα 18 γενικά μουσικά είδη είναι: Ambience, Blues, Classical, Country, Electronica, Folk, Hard, Hip Hop, Jazz, New Age, Pop, Reggae, Rhythm& Blues, Rock, Rock& Roll, Soul, Variety, και World.

Σύνολο Δεδομένων Τζανετάκη

Οι Τζανετάκης και Cook δημιούργησαν το 2002 μία βάση δύο επιπέδων για πειράματα αναγνώρισης μουσικού είδους (GTZAN dataset) [126]. Στο πρώτο επίπεδο υπάρχουν 10 βασικά μουσικά είδη, ενώ το δεύτερο επίπεδο περιέχει 4 υπο-είδη για την κατηγορία Classical και 6 υπο-είδη για την κατηγορία Jazz. Για κάθε μουσικό είδος και υπο-είδος δημιουργήθηκαν 100 αρχεία, διάρκειας 30 sec. το καθένα. Να σημειωθεί ότι τα πειράματα γενικής ταξινόμησης που πραγματοποίησε ο Τζανετάκης στο [126] έγιναν ξεχωριστά από τα πειράματα ταξινόμησης στα υπο-είδη της κλασσικής και jazz μουσικής. Πειράματα στην βάση του Τζανετάκη πραγματοποιήθηκαν από τον ίδιο στο [126], από τον Li στο [76], από τον Lidy στο [78], τον Bergstra στο [12],τον Holzapfel στο [56], τον Μπενέτο [8, 9] και πραγματοποιούνται και στην **παρούσα διατριβή**.

Παρόλου που τα ονόματα των καλλιτεχνών που αντιστοιχούν σε κάθε τραγούδι δεν είναι γνωστά, από το άκουσμα της βάσης, συνάγεται ότι δεν υπάρχουν περισσότερα από ένα μουσικό κομμάτι ανά καλλιτέχνη. Επομένως το *φαινόμενο του παραγωγού* [97] (producer effect) δεν εμφανίζεται στο συγκεκριμένο σύνολο δεδομένων, γεγονός που καθιστά τη βάση ιδανική για την αξιολόγηση αλγορίθμων αναγνώρισης μουσικού είδους.

Σύνολο Δεδομένων ISMIR

Για το συνέδριο ISMIR 2004[4] δημιουργήθηκαν 2 σύνολα δεδομένων, για χρήση σε διαγωνισμό αναγνώρισης μουσικού είδους. Το πρώτο σύνολο ονομάζεται ISMIRgenre και περιέχει 1458 αρχεία με γενικές κλάσεις μουσικών ειδών, καλύπτοντας 6 γενικές κλάσεις. Τα είδη που καλύπτει η βάση ISMIRgenre παρουσιάζονται στον Πίνακα 4.1. Πειράματα και στην βάση ISMIRgenre

[4] `http://ismir2004.ismir.net/ISMIR_Contest.html`

πραγματοποιήθηκαν στα πλαίσια του διαγωνισμού ISMIR 2004 από τους Lidy, Ellis, West, Pampalk, και Τζανετάκη [78] καθώς και στην **παρούσα διατριβή**.

Αντίστοιχη βάση δημιουργήθηκε για το συνέδριο ISMIR 2005[5], όπου δημιουργήθηκαν δύο βάσεις δεδομένων από διαφορετικές πηγές. Η πρώτη βάση αποτελείται από 1.515 κομμάτια που καλύπτουν 10 μουσικά είδη: classical, ambient, electronic, new-age, rock, punk, jazz, blues, folk, και ethnic. Η δεύτερη βάση αποτελείται από 1.414 κομμάτια που καλύπτουν 6 μουσικά είδη: rock, hip-hop, country, new-age, και reggae. Συγκεντρωτικά αποτελέσματα για τα δύο σύνολα δεδομένων του διαγωνισμού παρατίθενται στο [114].

Το 2007 στα πλαίσια του ISMIR-MIREX διεξήχθη πάλι διαγωνισμός για την αξιολόγηση αλγορίθμων αναγνώρισης μουσικού είδους. Η βάση που χρησιμοποιήθηκε περιελάμβανε 7000 μονοφωνικά μουσικά κομάτια διάρκειας 30 sec. το καθένα με συχνότητα δειγματοληψίας 22.05kHz σε αρχεία τύπου wav ομοιόμορφα κατανεμημένα σε 10 μουσικά είδη. Το περιεχόμενο της βάσης συνοψίζεται στον Πίνακα 4.2.

Μουσικό Είδος	**Αριθμός Αρχείων**
Classical	640
Electronic	229
Jazz - Blues	52
Metal - Punk	90
Rock - Pop	203
World	244

Πίνακας 4.1: Τα μουσικά είδη που καλύπτονται από τη βάση ISMIRgenre.

Σύνολα Δεδομένων Meng

Το 2005 ο Meng δημιούργησε δύο σύνολα δεδομένων για πειράματα σε αναγνώριση μουσικού είδους [90]. Το πρώτο σύνολο περιείχε 100 κομμάτια, οργανωμένα σε 5 είδη: classical, hard rock, jazz, pop, και techno. Το δεύτερο σύνολο δεδομένων αποτελείται από 354 δείγματα διάρκειας 30 sec το κάθε ένα, προερχόμενο από δωρεάν δείγματα που παρέχει το Amazon. Το δεύτερο σύνολο οργανώνεται σε 6 μουσικά είδη: classical, country, jazz, rap, rock, και techno.

Το 2008 ο ίδιος πρότεινε το σύνολο δεδομένων Garageband [91] το οποίο περίεχει 16706 μουσικά κομματια που ανήκουν στα ακόλουθα 47 μουσικά έιδη και στυλ: Acoustic, Alternative Metal, Alternative Pop, Alternative Rock, Ambient, Americana, Blues, Blues Rock, Classical, Comedy, Country, Dance Electronic, Electronica, Emo, Experimental Electronica, Experimental Rock, Folk, Folk Rock, Funk, Groove Rock, Hard Rock, Hardcore Metal, Hip

[5] `http://www.music-ir.org/evaluation/mirex-results/audio-genre/index.html`

Μουσικό Είδος	Αριθμός Αρχείων
Blues	700
Jazz	700
Country/Western	700
Baroque	700
Classical	700
Romantic	700
Electronica	700
Hip-Hop	700
Rock	700
HardRock/Metal	700

Πίνακας 4.2: Τα μουσικά είδη που καλύπτονται από τη βάση MIREX2007.

Hop, Indie Rock, Industrial, Instrumental Rock, Jazz, Latin, Metal, Modern Rock, Pop, Pop Punk, Pop Rock, Power Pop, Progressive Rock, Punk, R&B, Rap, Reggae, Rock, Ska, Spoken Word, Techno, Trance, World, WorldFusion. Η βάση Garageband οργανώνεται σε μια ιεραρχία δυο επιπέδων οπου στο πρώτο επίπεδο τοποθεντούνται τα 18 μουσικά έιδη τα οποια αποτελούνται απο μουσικά στυλ τοποθετημένα σε ένα δευτερο επίπεδο. Η ιεραρχεία συνοψίζεται στον Πίνακα 4.3.

Το σύνολο δεδομένων διατίθεται για ερευνητικους σκοπους. Προς το παρόν δεν έχει χρησιμοποιθεί σε πειράματα αναγνώρισης μουσικού έιδους.

Σύνολο Δεδομένων USPOP

Η βάση USPOP2 αποτελείται από ολόκληρα τραγούδια, συνοδευμένα από την AllMusic μεταπληροφορία τους. Σε αντίθεση με τις άλλες βάσεις δεδομένων η βάση USPOP2 δεν διατίθεται σε μορφή audio. Ο Dan Ellis διαθέτει τη βάση ως προϋπολογισμένα Mel-scale Frequency Cepstral Coefficients-MFCCs.

Τα τραγούδια της USPOP επιλέχθηκαν έτσι ώστε να αντιπροσωπεύουν την pop μουσική. Η βάση δεδομένων περιέχει μουσικά κομμάτια από 400 καλλιτέχνες. Η κατηγοριοποίηση των τραγουδιών σε μουσικά είδη έγινε με βάση τα αποτέλεσμα των επερωτήσεων στο AllMusic. Περισσότερες πληροφορίες για την βάση δεδομένων μπορούν να αναζητηθούν στην ιστοσελίδα της USPOP[6].

[6] `http://labrosa.ee.columbia.edu/projects/musicsim/uspop2002.html`

Μουσικό Είδος	Μουσικό Στυλ - Υποείδος
Rock	Alternative pop, Pop, Pop rock, Power pop,Alternative Rock, Indie Rock, Hard Rock, Modern Rock,Rock
Progressive Rock	Instrumental Rock, Progressive Rock
Folk/Country	Acoustic, Folk, Folk Rock, Americana, Country
Punk	Emo,Pop Punk, Punk
Heavy Metal	Alternative Metal, Hardcore Metal, Metal
Funk	Funk, Groove Rock, R& B
Jazz	Jazz
Electronica	Ambient, Electronica, Electronic, Experimental Electronica,Experimental Rock
Latin	Latin,World, World Fusion
Classical	Classical
Techno	Dance,Techno,Trance
Industrial	Industrial
Blues	Blues, Blues Rock
Reggae	Reggae
Ska	Ska
Comedy	Comedy
Rap	Hip-Hop, Rap
Spoken Word	Spoken Word

Πίνακας 4.3: Σύνολο δεδομένων Garageband, οργανωμένο σε είδη και στυλ.

Σύνολο Δεδομένων Homburg

Το 2005 ο Homburg [57] δημιούργησε ένα σύνολο δεδομένων με σκοπό να το χρησιμοποιήσει σε πειράματα ταξινόμησης και ομαδοποίησης μουσικής. Το σύνολο δεδομένων αποτελείται 1886 μουσικά κομμάτια ανομοιογενώς κατανεμημένα σε 9 μουσικά είδη. Κάθε αρχείο είναι σε μορφή mp3 με συχνότητα δειγματοληψίας 44.1 kHz και διάρκεια 10 sec. Μεταπληροφορίες για κάθε μουσικό κομμάτι είναι επίσης διαθέσιμες. Η βάση έχει δυο σοβαρά μειονεκτήματα, το πρώτο έγκειται στη μικρή χρονική διάρκεια των τραγουδιών, το δεύτερο είναι ότι οι ετικέτες δεν αντιπροσωπεύουν μόνο μουσικά είδη αλλά και μουσικά στυλ. Για παράδειγμα, το στυλ του Alternative αποτελεί υποείδος του μουσικού είδους Rock, παρόλα αυτά στη βάση θεωρούνται

ως δυο διαφορετικά μουσικά είδη. Τα μειονεκτήματα αυτά έχουν ως συνέπεια ότι η βάση δεν χρησιμοποιήθηκε σε καμία εργασία για την αναγνώριση μουσικού είδους, παρόλο που διατίθεται δωρεάν μέσω της ιστοσελίδας της, στο πανεπιστημίο του Dortmunt. Το περιεχόμενο της βάσης συνοψίζεται στον Πίνακα 4.4.

Μουσικό Είδος	**Αριθμός Αρχείων**
Blues	120
Electronic	113
Jazz	319
Pop	116
Rap/HipHop	300
Rock	504
Folk/Country	222
Alternative	145
Funk/Soul	47

Πίνακας 4.4: Τα μουσικά είδη που καλύπτονται από τη βάση του Homburg

4.2.2 Χαρακτηριστικά Περιγραφής Μουσικού Είδους

Το πρόβλημα της αυτόματης αναγνώρισης μουσικού είδους είναι εκ φύσεως πρόβλημα αναγνώρισης προτύπων. Ως εκ τούτου η επίλυση του απαιτεί την εξαγωγή χαρακτηριστικών που επιτρέπουν την περιγραφή του μουσικού είδους. Τέτοιου είδους χαρακτηριστικά περιγραφής που να καλύπτουν όλα (ή τα περισσότερα) μουσικά είδη, ανά τον χρόνο και τον τόπο μπορούν να προκύψουν από την ψηφιακή κυματομορφή του αρχείου μουσικής. Όμως, τα δείγματα της κυματομορφής δεν μπορούν να χρησιμοποιηθούν απευθείας σε εφαρμογές αναγνώρισης, λόγω του μεγάλου όγκου πληροφορίας που περιέχουν. Επιπλέον, η πληροφορία που εμπεριέχεται στα δείγματα της κυματομορφής είναι χαμηλού επιπέδου και δεν μπορεί να χρησιμοποιηθεί για σημασιολογική αναπαράσταση ενός μουσικού είδους. Σαν αποτέλεσμα, το πρώτο βήμα στην αναγνώριση μουσικού είδους είναι η εξαγωγή χαρακτηριστικών μεσαίου και υψηλού επιπέδου από την ψηφιακή επεξεργασία της κυμματομορφής.

Ο Pachet χωρίζει τα χαρακτηριστικά που χρησιμοποιούνται στην περιγραφή μουσικού είδους σε 3 κατηγορίες [96]:

- Χαρακτηριστικά χροιάς (timbre)
- Χρονορυθμικά χαρακτηριστικά
- Χαρακτηριστικά βασισμένα στη θεμελιώδη συχνότητα (pitch).

Αντίστοιχα, ο Scaringella πραγματοποιεί μία παρόμοια οργάνωση των χαρακτηριστικών, χωρίζοντάς τα σε 3 κατηγορίες [114]:

- Χαρακτηριστικά χροιάς (timbre)
- Ρυθμικά χαρακτηριστικά
- Μελωδικά-αρμονικά χαρακτηριστικά.

Τα παραπάνω χαρακτηριστικά χρησιμοποιούνται από την πλειονότητα των αλγορίθμων αυτόματης αναγνώρισης μουσικού είδους. Τα χαρακτηριστικά εξάγονται από κάθε μουσικό κομμάτι και συνδυάζονται σε ένα ενιαίο διάνυσα χαρακτηριστικών το οποίο αποτελεί την είσοδο σε κάποιο ταξινομητή. Η λογική αυτή συχνά αναφέρεται ως *ασκός χαρακτηριστικών* (bug of features) και στερείται θεωρητικής θεμελίωσης. Στην παρούσα διατριβή θα εγκαταλείψουμε την λογική αυτή της εξαγωγής χαρακτηριστικών και θα χρησιμοποιήσουμε χαρακτηριστικά που προκύπτουν από τις καλά θεμελιωμένες βιο-εμπνευσμένες αναπαραστάσεις που παρουσιάστηκαν στο Κεφάλαιο 2. Παρόλα αυτά αναφέρουμε στην ενότητα αυτή τα συνήθη χαρακτηριστικά περιγραφής μουσικού είδους για λόγους πληρότητας.

Χαρακτηριστικά χροιάς

Η χροιά ορίζεται στη βιβλιογραφία ως το χαρακτηριστικό που κάνει δύο ήχους στην ίδια θεμελιώδη συχνότητα και με την ίδια ένταση να εκλαμβάνονται ως διαφορετικοί [114]. Τα χαρακτηριστικά που περιγράφουν την χροιά εξετάζουν συνήθως τη φασματική κατανομή του σήματος, αν και μερικά υπολογίζονται στο πεδίο του χρόνου. Ο Peeters δημιούργησε μία εξαντλητική λίστα περιγραφής χαρακτηριστικών χροιάς στο [100].

Χρονικά χαρακτηριστικά

Υπολογίζονται για κάθε βραχυχρόνιο διάστημα (frame) του σήματος:

- Ρυθμός μηδενισμών (Zero-Crossing Rate): Υπολογίζει τον αριθμό των φορών που μηδενίζεται το σήμα σε ένα δίαστημα. Ένα ενθόρυβο σήμα έχει μεγάλη τιμή ρυθμού μηδενισμών. Χρησιμοποιήθηκε από τους Τζανετάκη [126], Meng [90], Burred [21], Li [76], και Cataltepe [23].
- Συντελεστές γραμμικής πρόβλεψης (Linear Prediction Coefficients): Χρησιμοποιούνται κυρίως στην επεξεργασία ομιλίας. Οι συντελεστές αναφέρονται σε ένα φίλτρο όλο-πόλων που θεωρείται ότι παράγει το ηχητικό σήμα έχοντας σαν είσοδο μία περιοδική διέγερση, που περιέχει πληροφορία για την θεμελιώδη συχνότητα. Χρησιμοποιήθηκαν από τους Τζανετάκη [126], Pachet [3], και Meng [90].

Χαρακτηριστικά ενέργειας

Αναφέρονται στο ενεργειακό περιεχόμενο του σήματος:

- Τετραγωνική ρίζα της ενέργειας (Root Mean Square Energy): Είναι μέτρο της ισχύος ενός διαστήματος. Χρησιμοποιήθηκε από τον Burred [21].

- Λόγος χαμηλής ενέργειας (Low Energy Rate): Ποσοστό των διαστημάτων ενός σήματος που έχουν τετραγωνική ρίζα της ενέργειας μικρότερη από τη μέση ενέργεια. Χρησιμοποιήθηκε από τους Τζανετάκη [126], Meng [90], Burred [21], Li [76], και Cataltepe [23].

Φασματικά χαρακτηριστικά

Περιγράφουν τη μορφή του φάσματος ισχύος ενός διαστήματος:

- Κέντρο βάρους του φάσματος (Spectral Centroid): ορίζεται ως το κέντρο βάρους του φάσματος ισχύος και είναι μέτρο του φασματικού σχήματος. Χρησιμοποιήθηκε από τους Τζανετάκη [126], Burred [21], Pachet [3], Li [76], και Cataltepe [23]. Παραλλαγές του αποτελούν η διασπορά του φάσματος (Spectral Spread) και το επίπεδο του φάσματος (Spectral Flatness), που χρησιμοποιήθηκαν από τον Burred [21].

- Συχνότητα απόσβεσης (Spectral Roll-off Frequency): Μετράει σε πόσο υψηλές συχνότητες (συνήθως 85%) συναντάται ένα συγκεκριμένο ποσοστό της ενέργειας του σήματος. Χρησιμοποιήθηκε από τους Pachet [3], Τζανετάκη [126], Li [76], Barbedo [6], και Cataltepe [23].

- Διακύμανση φάσματος (Spectrum Flux): ορίζεται ως η μέση διακύμανση του φάσματος μεταξύ δύο γειτονικών διαστημάτων. Χρησιμοποιήθηκε από τους Τζανετάκη [126], Burred [21], Pachet [3], Li [76], Barbedo [6], και Cataltepe [23].

- Mel-χασματικοί συντελεστές (Mel Frequency Cepstral Coefficients): Υπολογίζονται από τους συντελεστές βραχυχρόνιου διακριτού μετασχηματισμού Fourier, οι οποίοι φιλτράρονται από μία τράπεζα φίλτρων. Χρησιμοποιήθηκαν από τους Τζανετάκη [126], Burred [21], Meng [90], Pachet [3], Li [76], και Cataltepe [23].

Αντιλαμβανόμενα χαρακτηριστικά

Τα αντιλαμβανόμενα χαρακτηριστικά (perceptual features) υπολογίζονται χρησιμοποιώντας ένα μοντέλο της ανθρώπινης ακοής:

- Αντιλαμβανόμενη ένταση (Perceptual Loudness): χρησιμοποιεί την ένταση ενός διαστήματος ήχου μετασχηματισμένη με βάση ένα μοντέλο ακοής. Χρησιμοποιήθηκε από τον Burred [21] και τον Barbedo [6].

Χρονορυθμικά χαρακτηριστικά

Ο Pachet υποστηρίζει ότι ένα σύστημα αυτόματης αναγνώρισης μουσικού είδους δεν πρέπει να χρησιμοποιεί μόνο χαρακτηριστικά χροιάς, αλλά να χρησιμοποιεί και ρυθμική πληροφορία [3]. Ακριβής ορισμός του ρυθμού δεν υπάρχει, αλλά αναφέρεται στην έννοια της χρονικής επανάληψης [114]. Διαισθητικά, είναι προφανές ότι το ρυθμικό περιεχόμενο ενός κομματιού μπορεί να χρησιμοποιηθεί για διαχωρισμό έντονα ρυθμικών κομματιών (πχ. rock) και κομματιών χωρίς έντονη αίσθηση ρυθμού (πχ. κλασική μουσική).

Ο Gouyon εξέτασε τα προτεινόμενα ρυθμικά χαρακτηριστικά στη βιβλιογραφία [48]. Υπάρχουν πολλές διαφορετικές όψεις στην αναζήτηση ρυθμού: εξαγωγή tempo, αναζήτηση κρούσης (beat), αναγωγή μέτρου. Συνήθως οι μέθοδοι εξαγωγής ρυθμικών χαρακτηριστικών αναζητούν περιοδικότητες στο σήμα, χρησιμοποιώντας συνήθως συναρτήσεις αυτοσυσχέτισης ή τον μετασχηματισμό Fourier του σήματος.

Ο Τζανετάκης προτείνει τη χρήση ενός ιστογράμματος κρούσης (beat histogram), το οποίο δημιουργείται από την συνάρτηση αυτοσυσχέτισης του σήματος [126]. Παρατηρώντας τα βάρη στις διάφορες περιοδικότητες του σήματος, εξάγεται μία ρυθμική 'ποσότητα', που συντελεί στο διαχωρισμό μουσικών κομματιών με έντονο ρυθμό σε σχέση με αυτά που δεν έχουν δυνατή την αίσθηση του ρυθμού. Ένα παρόμοιο ιστόγραμμα χρησιμοποιείται από τον Burred, από το οποίο εξάγονται δύο ρυθμικά χαρακτηριστικά: η δύναμη κρούσης (beat strength) και ρυθμική κανονικότητα (rhythmic regularity) [21]. Ο Li χρησιμοποιεί επίσης το ιστόγραμμα κρούσης [76]. Το υπολογίζει εφαρμόζοντας τη συνάρτηση αυτοσυσχέτισης στην περιβάλλουσα του σήματος. Τα ρυθμικά χαρακτηριστικά που εξάγονται είναι: το σχετικό πλάτος των δύο κορυφών του ιστογράμματος, ο λόγος του πλάτους των δύο κορυφών, οι περίοδοι των δύο κορυφών, και το συνολικό άθροισμα του ιστογράμματος. Ο Meng, εκτός από το ιστόγραμμα κρούσης, προτείνει και τη χρήση φάσματος κρούσης (beat spectrum) το οποίο χρησιμοποιεί τους Mel χασματικούς συντελεστές για τον υπολογισμό του [90]. Ο Lidy προτείνει τα χαρακτηριστικά ρυθμικών προτύπων (rhythm pattern features) [78]. Τα χαρακτηριστικά προκύπτουν εφαρμόζοντας μετασχηματισμό Fourier και διαφορικό φίλτρο στους συντελεστές συγκεκριμένης αντιλαμβανόμενης έντασης (specific loudness sensation coefficients). Επίσης χρησιμοποιείται το ιστόγραμμα ρυθμού (rhythm histogram), το οποίο περιέχει πληροφορία ρυθμού ανά συχνοτική μπάντα. Τέλος, ο Gouyon χρησιμοποιεί μία πλειάδα χαρακτηριστικών περιγραφής ρυθμού [48]. Χρησιμοποιώντας χαρακτηριστικά περιγραφής tempo, περιοδικότητας, ισχύος περιοδικότητας, κέντρου βάρους

ιστογράμματος περιοδικότητας, χαρακτηριστικά περιγραφής κρουστότητας (percussiveness), και χαρακτηριστικά περιγραφής του ιστογράμματος παύσεων (interval histogram).

Αρμονικά και μελωδικά χαρακτηριστικά

Η αρμονία μπορεί να οριστεί ως η χρήση του μουσικού τόνου μαζί με τις αρμονικές (συγχορδίες στην περίπτωση της δυτικής μουσικής). Αντιθέτως, ως μελωδία ορίζουμε μία ακολουθία θεμελιωδών συχνοτήτων. Ενώ η αρμονία αναφέρεται συνήθως ως το 'κάθετο' στοιχείο της μουσικής, η μελωδία αναφέρεται ως το 'οριζόντιο' στοιχείο της. Η μελωδική και αρμονική ανάλυση χρησιμοποιείται αιώνες από μουσικολόγους για τη μελέτη μουσικών δομών και η χρήση της στο πρόβλημα της αναγνώρισης μουσικού είδους μπορεί να οδηγήσει σε θετικά αποτελέσματα. Ο Gómez κάνει μία ανασκόπηση της περιγραφής και εξαγωγής μελωδίας από ηχητικό σήμα στο [46].

Ο Τζανετάκης το 2002 αποπειράθηκε πρώτος να χρησιμοποιήσει χαρακτηριστικά βασισμένα στον μουσικό τόνο για αναγνώριση μουσικού είδους [126]. Το διάνυσμα χαρακτηριστικών μουσικού τόνου (pitch content feature set) που προτείνει βασίζεται σε πολλαπλές τεχνικές ανίχνευσης μουσικού τόνου. Στην τεχνική που προτείνεται, το σήμα αποσυντίθεται σε δύο συχνοτικές μπάντες, πάνω και κάτω των 1000 Hz. Υπολογίζονται στην συνέχεια οι περιβάλλουσες για κάθε μπάντα. Αθροίζονται οι περιβάλλουσες και υπολογίζεται η αυτοσυσχέτιση του αποτελέσματος. Οι κορυφές της αυτοσυσχέτισης αντιστοιχούν στους μουσικούς τόνους του σήματος. Στη συνέχεια δημιουργείται ένα ιστόγραμμα μουσικού τόνου (pitch histogram) με βάση τις κορυφές. Οι συχνότητες που αντιστοιχούν σε κάθε κορυφή του ιστογράμματος μετατρέπονται σε μουσικές νότες και ονομάζονται σύμφωνα με το πρωτόκολλο MIDI. Δημιουργούνται δύο ιστογράμματα, το διπλωμένο (folded), που έχει υποστεί την μετατροπή σε MIDI, και το απλωμένο (unfolded), που δεν έχει υποστεί την μετατροπή. Τα τελικά χαρακτηριστικά που προκύπτουν από το ιστόγραμμα μουσικού τόνου είναι:

- FA0: πλάτος της μέγιστης κορυφής του διπλωμένου (folded) ιστογράμματος, αντιστοιχώντας στην τονική κλίμακα του δείγματος.
- UP0: Περίοδος της μέγιστης κορυφής στο απλωμένο (unfolded) ιστόγραμμα, που αντιστοιχεί στο εύρος σε οκτάβες του έργου.
- FP0: Περίοδος της μέγιστης κορυφής του διπλωμένου ιστογράμματος.
- IPO1: Διάστημα μουσικών τόνων ανάμεσα στις δύο υψηλότερες κορυφές του διπλωμένου ιστογράμματος, δείχοντας το κύριο τονικό διάστημα που χρησιμοποιείται στο κομμάτι (στα απλά έργα ισχύει το διάστημα τονικής-δεσπόζουσας).

- SUM: το άθροισμα του ιστογράμματος.

Ο Cataltepe επίσης χρησιμοποιεί 5 χαρακτηριστικά βασισμένα στο ιστόγραμμα μουσικού τόνου [23].

4.2.3 Αλγόριθμοι Αναγνώρισης

Οι τεχνικές αναγνώρισης μουσικού είδους κατατάσσονται σε δύο κατηγορίες στη βιβλιογραφία: στις μεθόδους χωρίς επίβλεψη (όπου δεν δίνεται ιεραρχία, αλλά η ιεραρχία προκύπτει ομαδοποιώντας τα μουσικά κομμάτια) και στις μεθόδους με επίβλεψη (όπου δίνεται η ιεραρχία μουσικών ειδών, το σύστημα εκπαιδεύεται και στη συνέχεια ελέγχεται η απόδοση του συστήματος με τα δεδομένα ελέγχου).

Μέθοδοι χωρίς επίβλεψη

Στις μεθόδους χωρίς επίβλεψη, τα δεδομένα ομαδοποιούνται με βάση τα χαρακτηριστικά τους και τα μέτρα ομοιότητας που χρησιμοποιούνται, και προκύπτει η ταξινόμηση. Το πλεονέκτημα αυτών των μεθόδων είναι ότι δεν υπάρχει ο περιορισμός μίας δεδομένης ιεραρχίας, που μπορεί να περιέχει ασάφειες και επικαλύψεις. Επίσης, κάποια κομμάτια πολύ απλά μπορεί να μην ανήκουν σε κανένα από τα δοσμένα μουσικά είδη.

Κάθε μουσικό κομμάτι αναπαριστάται από ένα σύνολο χαρακτηριστικών, όπως παρουσιάστηκε στην ενότητα 4.2.2. Επιλέγεται κατόπιν ένα μέτρο ομοιότητας για να συγκρίνει τα κομμάτια μεταξύ τους. Οι αλγόριθμοι ομαδοποίησης χρησιμοποιούν τα μέτρα ομοιότητας για την οργάνωση των μουσικών κομματιών σε ομάδες που έχουν κοινά χαρακτηριστικά.

Το πιο απλό μέτρο ομοιότητας που χρησιμοποιείται για να μετρήσει την απόσταση ανάμεσα σε δύο διανύσματα χαρακτηριστικών είναι είτε η Ευκλείδεια απόσταση ή το μέτρο ομοιότητας συνημιτόνου (cosine similarity measure). Όμως, αυτά τα μέτρα ομοιότητας πρέπει να εφαρμόζονται μόνο όταν τα διανύσματα χαρακτηριστικών είναι χρονοαμετάβλητα. Αλλιώς, δύο παρόμοια μουσικά κομμάτια μπορεί να θεωρηθούν ανόμοια ως προς το μέτρο, αν τα χαρακτηριστικά τους μεταβληθούν με τον χρόνο. Για να δημιουργηθεί μία χρονοαμετάβλητη αναπαράσταση μιας χρονοσειράς διανυσμάτων χαρακτηριστικών, συνήθως δημιουργούνται στατιστικά μοντέλα της κατανομής των χαρακτηριστικών και έπειτα χρησιμοποιείται κάποιο μέτρο ομοιότητας για την σύγκριση των κατανομών.

Τυπικά μοντέλα είναι τα μοντέλα μείγματος Γκαουσιανών (Gaussian mixture models - GMMs). GMMs χρησιμοποιήθηκαν για την δημιουργία μοντέλων χροιάς στο [89]. Επίσης, η απόκλιση Kullback-Leibler χρησιμοποιείται για να υπολογίσει την απόσταση ανάμεσα σε κατανομές, αλλά δεν συστήνεται για GMMs. Άλλο μέτρο που χρησιμοποιείται για σύγκριση

κατανομών είναι η απόσταση Earth Movers, που χρησιμοποιήθηκε από τον Pampalk [97]. Ο Shao χρησιμοποίησε επίσης κρυμμένα μοντέλα Markov (hidden Markov models - HMMs), για να μοντελοποιηθεί η σχέση των χαρακτηριστικών στον χρόνο [118].

Όσον αφορά τους αλγορίθμους ομαδοποίησης που χρησιμοποιούνται, ο Shao χρησιμοποίησε ιεραρχικούς συσσωρευτικούς αλγορίθμους ομαδοποίησης (agglomerative hierarchical clustering) στο [118]. Οι ιεραρχικοί συσσωρευτικοί αλγόριθμοι αρχίζουν με N ομάδες (όπου N είναι ο αριθμός των κομματιών) και ενώνουν σταδιακά τις ομάδες χρησιμοποιώντας ένα μέτρο ομοιότητας.

Ο αυτο-οργανωνόμενος χάρτης (self-organizing map - SOM) και ο αυξανόμενος ιεραρχικός αυτο-οργανωνόμενος χάρτης (growing hierarchical self-organizing map - GHSOM) χρησιμοποιούνται για ομαδοποίηση δεδομένων. Τα δεδομένα αναπαριστώνται σε έναν δισδιάστατο χώρο, έτσι ώστε όμοια διανύσματα χαρακτηριστικών να απεικονίζονται κοντά. Τα SOMs είναι τεχνητά νευρωνικά δίκτυα χωρίς επίβλεψη που δημιουργούν αντιστοιχίες μεταξύ δεδομένων πολλών διαστάσεων σε χώρους χαμηλών διαστάσεων. Τα GHSOM αποτελούν ειδική περίπτωση των SOM, τα οποία χρησιμοποιούν μία ιερρχική δομή πολλών επιπέδων, όπου σε κάθε επίπεδο αντιστοιχεί ένα SOM. Ο Rauber χρησιμοποίησε GHSOMs για να αναπαραστήσει οπτικά μουσικές συλλογές στο [104]. Μία εφαρμογή ομαδοποίησης μουσικών κομματιών με ενδιαφέρον είναι ο 'Χάρτης του Mozart'[7], που δημιουργήθηκε από το Τεχνικό Πανεπιστήμιο Βιέννης. Στον χάρτη ομαδοποιούνται όλα τα έργα του W. A. Mozart σε 'νησιά' μουσικών ειδών, όπως συμφωνίες, σονάτες, κονσέρτα, τραγούδια, κ.α.

Το κύριο μειονέκτημα των μεθόδων χωρίς επίβλεψη είναι ότι οι προκύπτουσες ομάδες δεν ονοματίζονται. Γενικά όμως, οι ομάδες δεν απεικονίζουν απαραίτητα κάποιο μουσικό είδος, αλλά μουσικά κομμάτια με κάποια όμοια χαρακτηριστικά. Υποστηρίζεται ότι η έννοια του μουσικού είδους μπορεί να χαθεί σε βάθος χρόνου και να προκύψει μία οργάνωση των μουσικών δεδομένων με βάση την ομοιότητα στο [105].

Μέθοδοι με επίβλεψη

Οι μέθοδοι αναγνώρισης μουσικού είδους με επίβλεψη έχουν μελετηθεί περισσότερο στην βιβλιογραφία. Σ' αυτές τις τεχνικές υπάρχει μία βάση μουσικών κομματιών, η οποία πρέπει να αντιστοιχηθεί σε μία δοσμένη ιεραρχία, χρησιμοποιώντας αλγορίθμους μηχανικής μάθησης. Στο πρώτο βήμα, το σύστημα εκπαιδεύεται με δεδομένα για τα οποία είναι γνωστό το μουσικό είδος. Στο δεύτερο βήμα, ταξινομούνται αρχεία για τα οποία δεν είναι γνωστό το είδος (αρχεία ελέγχου). Παρακάτω περιγράφονται οι συνήθεις ταξινομητές που εφαρμόζονται στο πρόβλημα αναγνώρισης μουσικού είδους:

[7] `http://www.ifs.tuwien.ac.at/mir/mozart/`

- **Ταξινομητής K-πλησιέστερων γειτόνων** (K-nearest neighbor classifier - KNN): μη παραμετρικός ταξινομητής που βασίζεται στην ιδέα ότι ένας μικρός αριθμός γειτόνων καθορίζει την απόφαση για το δεδομένο ελέγχου. Δοσμένου ενός διανύσματος χαρακτηριστικών, επιλέγονται τα K πλησιέστερα διανύσματα σε αυτό. Το διάνυσμα ανήκει στην κλάση που συναντάται περισσότερο ανάμεσα στα K πλησιέστερα διανύσματα. Ο KNN αλγόριθμος χρησιμοποιήθηκε από τον Τζανετάκη [126] και τον Pampalk [97].

- **Μοντέλα μείγματος Γκαουσιανών** (Gaussian mixture models - GMMs): μοντελοποιούν την κατανομή των διανυσμάτων χαρακτηριστικών. Συνήθως χρησιμοποιείται ο αλγόριθμος μεγιστοποίησης και αναμενόμενης τιμής (expectation maximization - EM) για την εκτίμηση των παραμέτρων των Γκαουσιανών. Τα GMMs στο πεδίο της αναγνώρισης μουσικού είδους χρησιμοποιούνται συνήθως για τη δημιουργία μοντέλων χροιάς. Σαν ταξινομητές, μπορούν να χρησιμοποιηθούν σε συνδυασμό με ένα κριτήριο μέγιστης πιθανοφάνειας, για την εύρεση του μοντέλου που περισσότερο ταιριάζει στο δοσμένο κομμάτι ελέγχου. Ο Τζανετάκης χρησιμοποίησε GMMs για την μοντελοποίηση μουσικών ειδών [126]. Ο Burred χρησιμοποίησε μία δενδροειδή δομή από GMMs για τη μοντελοποίηση της μουσικής ιεραρχίας [21]. Τέλος, ο West χρησιμοποιεί έναν Γκαουσιανό ταξινομητή σε δενδροειδή δομή που χρησιμοποιεί την απόσταση Mahalanobis.

- **Κρυμμένα μοντέλα Markov** (hidden Markov models - HMMs): χρησιμοποιούνται κυρίως στην αναγνώριση ομιλίας, λόγω της ικανότητάς τους στο χειρισμό χρονοσειρών. Χρησιμοποιήθηκαν από τον Scaringella [111] για αναγνώριση σε 7 μουσικά είδη και από τον Soltau [121] για αναγνώριση σε 4 μουσικά είδη.

- **Ανάλυση Γραμμικού Διαχωρισμού** (linear discriminant analysis - LDA): η βασική ιδέα είναι η εύρεση ενός γραμμικού μετασχηματισμού που διαχωρίζει όσο το δυνατόν καλύτερα τις κλάσεις. Ο West χρησιμοποίησε LDA για να μειώσει τις διαστάσεις στο πρόβλημα της ταξινόμησης πριν να μοντελοποιήσει τα δεδομένα του με Γκαουσιανή κατανομή [129].

- **Μηχανές Εδραίων Διανυσμάτων** (support vector machines - SVMs): μαθαίνουν το υπερεπίπεδο μέγιστου περιθωρίου (maximum margin hyperplane), καθιστώντας τους ανθεκτικούς στην υπερπροσαρμογή (over-fitting). Χρησιμοποιήθηκαν από τον Scaringella [111] και τον Lidy [78] για πειράματα αναγνώρισης μουσικού είδους. Ο Mandel χρησιμοποίησε SVMs σε συνδυασμό με την απόσταση Kullback-Leibler [89].

- **Τεχνητά νευρωνικά δίκτυα** (artificial neural networks - ANN): χρησιμοποιούν σαν δομικά στοιχεία τεχνητούς νευρώνες για την επίλυση προβλημάτων. Τα πιο διαδεδομένα

δίκτυα είναι τα πολυστρωματικά perceptrons (multilayer perceptrons - MLPs), τα οποία μπορούν να προσεγγίσουν οποιαδήποτε μη γραμμική συνάρτηση. Ο Soltau χρησιμοποίησε μία παραλλαγή των MLPs για αναγνώριση μουσικού είδους, τα οποία περιέγραψε ως ρητή χρονική μοντελοποίηση με νευρωνικά δίκτυα (explicit time modeling with neural networks - ETM-NN) [121]. Στο ETM-NN, κάθε κρυμμένος νευρώνας μπορεί να θεωρηθεί ως ένα μουσικό γεγονός (μουσική φράση, ρυθμικό ή μελωδικό μοτίβο).

4.2.4 Συγκεντρωτικά Αποτελέσματα - Συμπεράσματα

Στην συνέχεια παρουσιάζουμε συγκεντρωτικά αποτελέσματα των πιο αξιόλογων αλγορίθμων αναγνώρισης μουσικού είδους. Το βασικό πρόβλημα των αποτελεσμάτων είναι ότι δεν είναι άμεσα συγκρίσιμα μεταξύ τους. Τούτο οφείλεται στο γεγονός ότι κατά την πειραματική διαδικασία δεν ακολουθείται κοινό πρωτόκολλο πειραμάτων. Για παράδειγμα σε κάποιες εργασίες αναφέρεται ως αποτέλεσμα μόνο το καλύτερο αποτέλεσμα από ένα σύνολο πειραμάτων, σε άλλες χρησιμοποιείται δεκαπλή διασταυρωμένη επικύρωση 10-fold cross validation ενώ σε άλλες πενταπλή διασταυρωμένη επικύρωση 5-fold cross validation. Στον Πίνακα 4.5 συνοψίζονται τα αποτελέσματα των καλύτερων αλγορίθμων αναγνώρισης μουσικού είδους που έχουν προταθεί στην βιβλιογραφία.

Αναφορά	**Σύνολο Δεδομένων**	**Ακρίβεια**
Bergstra *et al.* [12]	GTZAN	82.50%
Li *et al.* [76]	GTZAN	78.50%
Lidy *et al.* [79]	GTZAN	76.80%
Benetos *et al.* [9]	GTZAN	75.00%
Holzapfel *et al.* [56]	GTZAN	74.00%
Tzanetakis *et al.* [126]	GTZAN	61.00%
Holzapfel *et al.* [56]	ISMIR2004	83.50%
Pampalk *et al.* [97]	ISMIR2004	82.30%
Lidy *et al.* [78]	ISMIR2004	79.70%
Bergstra *et al.* [12]	MIREX2005	82.34%
Lidy *et al.* [79]	MIREX2007	75.57%
Mandel *et al.* [85]	MIREX2007	75.03%

Πίνακας 4.5: Ακρίβεια ορθής ταξινόμησης διάφορων αξιοσημείωτων αλγορίθμων αναγνώρισης μουσικού είδους.

Σε μια προσπάθεια ερμηνείας των **αποτελεσμάτων** που παρουσιάζονται στον Πίνακα 4.5, παρατηρούμε ότι η ακρίβεια ορθής ταξινόμησης μουσικού είδους των συστημάτων δεν έχει καταφέρει να ξεπεράσει ουσιαστικά το 83%. Το γεγονός αυτό μπορεί να οφείλεται αφενός στην ποιότητα των δεδομένων που χρησιμοποιούνται για την αξιολόγηση των αλγορίθμων και αφετέρου ίσως το ποσοστό αυτό να αποτελεί το άνω όριο διακριτοποίησης των μουσικών ειδών χρησιμοποιώντας πληροφορίες που εξάγονται μόνο από το μουσικό σήμα. Συμπεραίνουμε λοιπόν ότι απαιτείται ακόμη πιο εντατική προσπάθεια προς την κατεύθυνση οι αλγόριθμοι αναγνώρισης μουσικού είδους να αποκτήσουν πρακτικώς χρήσιμες επιδόσεις.

Κεφάλαιο 5

Πολυγραμμικές Τεχνικές Ανάλυσης Υποχώρων

5.1 Εισαγωγή

Οι *τεχνικές ανάλυσης υποχώρων* (subspace analysis techniques), αποκαλύπτουν δομές χαμηλών διαστάσεων σε χώρους πολλών διαστάσεων. Μία κατάλληλη αναπαράσταση των δεδομένων σε χώρο χαμηλών διαστάσεων μπορεί να οδηγήσει στην εύρεση χαρακτηριστικών γνωρισμάτων των δεδομένων, να κάνει πιο εύκολη την ερμηνεία τους, και να προσδιορίσει τους παράγοντες που συνθέτουν τα δεδομένα καθαυτά. Οι τεχνικές ανάλυσης υποχώρων έχουν χρησιμοποιηθεί εκτεταμένα και με επιτυχία ως μέθοδοι εξαγωγής χαρακτηριστικών σε μια πληθώρα προβλημάτων αναγνώρισης προτύπων.

Στο παρόν κεφάλαιο, παρουσιάζονται εν συντομία οι βασικές γραμμικές τεχνικές ανάλυσης υποχώρων, και με περισσότερες λεπτομέρειες οι αντίστοιχες πολυγραμμικές τεχνικές ανάλυσης υποχώρων οι οποίες θα χρησιμοποιηθούν στην διατριβή για την εξαγωγή χαρακτηριστικών διανυσμάτων από τις αναπαραστάσεις φλοιού του ήχου. Ιδιαίτερο βάρος δίνεται στην παραγοντοποιήση μη αρνητικών τανυστών, όπου παρουσιάζεται η πλειονότητα των αλγορίθμων που έχει προταθεί στην βιβλιογραφία. Στο τέλος του κεφαλαίου παρουσίαζονται δύο νέοι αλγόριθμοι παραγοντοποιήσης μη αρνητικών τανυστών, με θεωρητικά θεμελιωμένες ιδιότητες σύγκλισης.

5.2 Γραμμικές Τεχνικές Ανάλυσης Υποχώρων

Στη βιβλιογραφία έχουν προταθεί πολλές τεχνικές ανάλυσης υποχώρων. Η πιο διαδεδομένη είναι η ανάλυση πρωτευουσών συνιστωσών (principal component analysis - PCA), η οποία μετασχηματίζει τα δεδομένα σε ένα νέο σύστημα συντεταγμένων, έτσι ώστε να μεγιστοποιείται η

μεταβλητότητα στον κύριο άξονα (πρωτεύουσα συνιστώσα). Σκοπός της ανάλυσης ανεξάρτητων συνιστωσών (independent component analysis - ICA) είναι η ανάλυση ενός πολυμεταβλητού σήματος σε προσθετικές συνιστώσες, θεωρώντας σαν περιορισμό την στατιστική ανεξαρτησία των συνιστωσών [62]. Η ανάλυση παραγόντων (factor analysis) θεωρεί ένα γραμμικό μοντέλο, έτσι ώστε οι παρατηρούμενες μεταβλητές των δεδομένων να δημιουργούνται ως γραμμικός συνδυασμός μη παρατηρούμενων μεταβλητών, που ονομάζονται παράγοντες. Η λανθάνουσα σημασιολογική ανάλυση (latent semantic analysis - LSA) [36] και η πιθανοτική λανθάνουσα σημασιολογική ανάλυση (probabilistic latent semantic analysis - PLSA) [55] είναι στατιστικές τεχνικές που προσπαθούν να αντιστοιχίσουν διανύσματα πολλών διαστάσεων σε χώρο χαμηλών διαστάσεων και χρησιμοποιούνται κυρίως στα πεδία ανάκτησης πληροφορίας και επεξεργασίας φυσικής γλώσσας. Τέλος, η παραγοντοποίηση μη αρνητικών πινάκων (non-negative matrix factorization - NMF), βρίσκει την παραγοντοποίηση ενός πίνακα, με τον περιορισμό ότι τα στοιχεία του λαμβάνουν μη αρνητικές τιμές [74, 81, 61].

5.3 Πολυγραμμικές Τεχνικές Ανάλυσης Υποχώρων

Αντίστοιχα με τις γραμμικές τεχνικές ανάλυσης υποχώρων για πολυμεταβλητά δεδομένα, πρόσφατα έχουν προταθεί και τεχνικές ανάλυσης πολυδιάστατων πολυμεταβλητών δεδομένων (multidimensional multivariate data analysis techniques). Ενώ οι γραμμικές τεχνικές ανάλυσης υποχώρων για πολυμεταβλητά δεδομένα χρησιμοποιούν πίνακες και τεχνικές γραμμικής άλγεβρας για την πραγματοποίηση των υπολογισμών, οι τεχνικές ανάλυσης πολυδιάστατων πολυμεταβλητών δεδομένων χρησιμοποιούν τανυστές και τεχνικές πολυγραμμικής άλγεβρας για την πραγματοποίηση των υπολογισμών. Η ανάγκη για την δημιουργία των πολυδιάστατων τεχνικών προήλθε από τα πεδία της ψυχολογίας και χημείας, όπου έγιναν οι πρώτες απόπειρες ανάλυσης και ερμηνείας τρισδιάστατων δεδομένων. Το 1964 ο Tucker πρότεινε μία παραλλαγή της PCA για ανάλυση πραγματικών τρισδιάστατων πινάκων στο πεδίο της ψυχομετρίας. Το μοντέλο που αναπτύχθηκε ονομάστηκε Tucker3 model [127].

5.3.1 PARAFAC

Η πιο διαδεδομένη τεχνική ανάλυσης τρισδιάστατων δεδομένων είναι η PARAFAC (από το parallel factor analysis), η οποία ονομάζεται επίσης και CANDECOMP (από το canonical decomposition) [19]. Η PARAFAC προτάθηκε για τα πεδία της ψυχομετρίας (psychometrics) και της χημειομετρίας (chemometrics). Βασικός σκοπός της μεθόδου PARAFAC είναι η αποσύνθεση ενός τρισδιάστατου τανυστή $\mathcal{D} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ σε άθροισμα εξωτερικών γινομένων

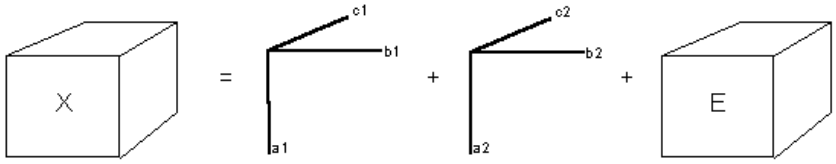
διανυσμάτων:

$$\mathcal{D} = \sum_{j=1}^{k} \mathbf{a}_j \otimes \mathbf{b}_j \otimes \mathbf{c}_j. \tag{5.1}$$

Η αποσύνθεση πραγματοποιείται με αλγόριθμο εναλασσόμενων ελαχίστων τετραγώνων (alternating least squares - ALS). Στην PARAFAC λύνεται το ακόλουθο πρόβλημα βελτιστοποίησης:

$$\min_{\mathbf{a},\mathbf{b},\mathbf{c}} \sum_{i_1 i_2 i_3} \left|\left| \mathcal{D}_{i_1 i_2 i_3} - \sum_{j=1}^{k} a_{j(i_1)} \otimes b_{j(i_2)} \otimes c_{j(i_3)} \right|\right|^2. \tag{5.2}$$

Βασικό πλεονέκτημα της PARAFAC είναι ότι βρίσκει μοναδική λύση με κατάλληλη επιλογή του k. Στο Σχήμα 5.1 απεικονίζεται η μοντελοποιήση που συνεπάγεται η PARAFAC. Στο συγκεκριμένο μοντέλο εισάγεται και θόρυβος, για να περιγράψει τα λάθη υπολογισμού, ο οποίος πρέπει να ελαχιστοποιηθεί.



Σχήμα 5.1: Αναπαράσταση της PARAFAC για $k = 2$ (από το [19]).

5.3.2 Πολυγραμμική Ανάλυση Πρωτευουσών Συνιστωσών - MPCA

Ο Kroonenberg έθεσε πρώτος τα θεμέλια για έναν αλγόριθμο που να αποτελεί πολυγραμμική επέκταση της ανάλυσης πρωτευουσών συνιστωσών [69]. Το μοντέλο τριών διαστάσεων που προτείνει είναι βασισμένο στο μοντέλο του Tucker:

$$\mathcal{D}_{i_1 i_2 i_3} = \sum_{k=1}^{K} \sum_{l=1}^{L} \sum_{m=1}^{M} g_{i_1 k} h_{i_2 l} e_{i_3 m} c_{klm}. \tag{5.3}$$

όπου $\mathcal{D} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ και $\mathcal{C} \in \mathbb{R}^{K \times L \times M}$. Το μοντέλο μπορεί εναλλακτικά να γραφεί ως:

$$\mathcal{D} = \mathcal{C} \times_1 \mathbf{G} \times_2 \mathbf{H} \times_3 \mathbf{E}. \tag{5.4}$$

όπου $\mathbf{G} \in \mathbb{R}^{I_1 \times K}$, $\mathbf{H} \in \mathbb{R}^{I_2 \times L}$, και $\mathbf{E} \in \mathbb{R}^{I_3 \times M}$. Για την επίλυση του προβλήματος χρησιμοποιείται η μέθοδος εναλλασσόμενων ελαχίστων τετραγώνων (alternating least squares - ALS).

Το 1986, ο Kapteyn κάνει μία επέκταση του μοντέλου 3-mode components analysis που πρώτος είχε προτείνει ο Tucker, σε N διαστάσεις [64]. Για την επίλυση του προβλήματος χρησιμοποιείται η μεθοδος ALS.

Πρόσφατα, ο Lu στο [80] προτείνε την *Πολυγραμμική Ανάλυση Πρωτευουσών Συνιστωσών* (Multilinear Principal Component Analysis - MPCA) ως μια γενίκευση της PCA για τανυστές N τάξης κατ᾽ αναλογία με την PCA. Έστω $\{\mathcal{D}_m \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}, m = 1, 2, \ldots, M\}$ ένα σύνολο M δειγμάτων που αναπαρίστανται ως τανυστές. Ο τανυστής διασποράς ορίζεται ως εξής:

$$\Psi_{\mathcal{D}} = \sum_{m=1}^{M} ||\mathcal{D}_m - \hat{\mathcal{D}}||_F^2. \tag{5.5}$$

όπου $\hat{\mathcal{D}}$ είναι ο μέσος (mean) τανυστής. Ο αντικειμενικός σκοπός της MPCA είναι να βρεθεί ένας πολυγραμμικός μετασχηματισμός τέτοιος ώστε να απεικονίζει τον αρχικό τανυστικό χώρο $\mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ σε έναν τανυστικό υποχώρο, έστω $\mathbb{R}^{P_1 \times P_2 \times \ldots \times P_N}$ με $P_n < I_n$, $\forall n$, τέτοιο ώστε ο υποχώρος να συλλαμβάνει την μέγιστη δυνατή μεταβλητότητα του αρχικού συνόλου τανυστών, υποθέτοντας ότι αυτή μεταβλητότητα μετριέται από την (5.5). Εφόσον $P_n < I_n$ η MPCA επιτυγχάνει μείωση των διαστάσεων.

Ο αλγόριμος υπολογισμού της MPCA έχει ως εξής:

Βήμα 1 (Προεπεξεργασία): Αφαίρεσε το μέσο τανυστή από του τανυστές εισόδου ως εξής: $\{\hat{\mathcal{D}_m} = \mathcal{D}_m - \hat{\mathcal{D}},\ m = 1, 2,\ 3, \ldots, M\}$ όπου $\hat{\mathcal{D}}$ είναι ο μέσος τανυστής.

Βήμα 2 (Αρχικοποίηση): Εφάρμοσε ιδιοανάλυση (eigenanalysis) στον $\mathbf{\Phi}^{(n)*} = \sum_{m=1}^{M} \hat{\mathbf{D}}_{m(n)} \hat{\mathbf{D}}_{m(n)}^T$ και όρισε το $\tilde{\mathbf{U}}^{(n)}$ έτσι ώστε να περιέχει τα ιδιοδιανύσματα που αντιστοιχούν στις $P_n\ \forall\ n = 1, 2, \ldots N.$ πιο σηματικές ιδιοτιμές.

Βήμα 3 (Τοπική Βελτιστοποιήση):

- Υπολόγισε $\{\tilde{\mathcal{Y}}_m = \tilde{\mathcal{D}}_m \times_1 \mathbf{U}^{(1)^T} \times_2 \mathbf{U}^{(2)^T} \times_3 \ldots \times_N \mathbf{U}^{(N)^T},\ m = 1, 2, \ldots, M\}$
- Υπολόγισε $\mathbf{\Psi}_{\mathcal{Y}_0} = \sum_{m=1}^{M} \|\tilde{\mathcal{Y}}_m\|_F^2$.
- For $k = 1 : K$

  For $n = 1 : N$

  - όρισε το $\tilde{\mathbf{U}}^{(n)}$ ετσι ώστε να περιέχει τα ιδιοδιανύσματα που αντιστοιχούν στις πιο σηματικές $P_n \forall n = 1, 2, \ldots N.$ ιδιοτιμές του πίνακα $\mathbf{\Phi}^{(n)}$.
  - Υπολόγισε τον $\tilde{\mathcal{Y}}_m, m = 1, 2, \ldots, M$ και τον $\mathbf{\Psi}_{\mathcal{Y}_k}$

  if $\|\mathbf{\Psi}_{\mathcal{Y}_k} - \mathbf{\Psi}_{\mathcal{Y}_{k-1}}\|_F < \varepsilon$ break; και goto Βήμα 4.

Βήμα 4 (Προβολή): Ο μικρότερων διαστάσεων τανυστής χαρακτηριστικών υπολογίζεται ως εξής:$\{\mathcal{Y}_m = \mathcal{D}_m \times_1 \tilde{\mathbf{U}}^{(1)^T} \times_2 \tilde{\mathbf{U}}^{(2)^T} \times_3 \ldots \times_N \tilde{\mathbf{U}}^{(N)^T},\ m = 1, 2, \ldots, M\}$

Η MPCA χρησιμοποιήθηκε από τον Lu στο πρόβλημα της αναγνώρισης βηματισμού (gait recognition) [80]. Στη παρούσα διατριβή χρησιμοποιείται ως πολυγραμμική μέθοδος εξαγωγής χαρακτηριστικών από τις φλοιώδεις αναπαραστάσεις του ήχου.

5.3.3 Υψηλής Τάξης Αποσύνθεση Ιδιαζουσών Τιμών - HOSVD

Ο Lathauwer προτείνει μία μέθοδο πολυγραμμικής αποσύνθεσης ιδιαζουσών τιμών (singular value decomposition -SVD) [72] [73]. Ονομάζει το μοντέλο ως SVD υψηλής τάξης (higher-order SVD - HOSVD) και είναι το ακόλουθο:

$$\mathcal{D} = \mathcal{S} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \times_3 \cdots \times_N \mathbf{U}^{(N)}. \tag{5.6}$$

όπου οι τανυστές $\mathcal{D}, \mathcal{S} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ και οι πίνακες $\mathbf{U}^{(i)} \in \mathbb{R}^{I_i \times I_i}$. Οι πίνακες $\mathbf{U}^{(i)}$ είναι ορθογώνιοι. Ο κεντρικός τανυστής $\mathcal{S}$ είναι όλο-ορθογώνιος (βλ. Ενότητα 3.3.7) και υπάρχει διάταξη των υπο-τανυστών του $\mathcal{S}$, σε αναλογία με την ανάλυση ιδιαζουσών τιμών:

$$||\mathcal{S}_{i_n=1}|| \geqslant ||\mathcal{S}_{i_n=2}|| \geqslant \cdots \geqslant ||\mathcal{S}_{i_n=I_n}|| \geqslant 0 \tag{5.7}$$

Ο Lathauwer αποδεικνύει ότι για την εύρεση του HOSVD απαιτείται υπολογισμός SVD στα αναπτύγματα του $\mathcal{D}$, ενώ ο κεντρικός τανυστής $\mathcal{S}$ υπολογίζεται στην συνέχεια:

$$\mathcal{S} = \mathcal{D} \times_1 \mathbf{U}^{(1)^T} \times_2 \mathbf{U}^{(2)^T} \times_3 \cdots \times_N \mathbf{U}^{(N)^T}. \tag{5.8}$$

Αποδεικνύεται ότι η λύση που δίνει η HOSVD είναι μοναδική. Επίσης, για $N = 2$ ο αλγόριθμος είναι ίδιος με τον SVD. Η HOSVD μπορεί να χρησιμοποιηθεί ακόμα για τον υπολογισμό του n-βαθμού του $\mathcal{D}$.

5.3.4 Παραγοντοποίηση Μη-αρνητικών Τανυστών - NTF

Ορισμός Προβλήματος

Η Παραγοντοποιήση Μη-αρνητικών Τανυστών αποτελεί μια κατηγορία πολυγραμμικών τεχνικών ανάλυσης υποχώρων οι όποιες παρέχουν σημαντικούς λανθάνοντες παράγοντες ή ουσιώδη χαρακτηριστικά τανυστών N-τάξης [32].

Δεδομένου ενος μη αρνητικού τανυστή N-τάξης έστω $\mathcal{D} \in \mathbb{R}_+^{I_1 \times \ldots \times I_N}$, $I_i \in \mathbb{Z}$, $i = 1, 2, \ldots, N$, και ενός θετικού ακεραίου k, ο NTF δύναται να προσεγγίσει τον τανυστή $\mathcal{D}$ ως ένα άθροισμα k τανυστών πρώτης τάξης.

$$\mathcal{D} \simeq \sum_{j=1}^{k} \mathbf{u}_1^j \otimes \mathbf{u}_2^j \otimes \cdots \otimes \mathbf{u}_N^j \tag{5.9}$$

όπου $\mathbf{u}_i^j \in \mathbb{R}_+^{I_i}$ Η παραγοντοποίηση μη αρνητικών πινάκων (Non-negative Matrix Factorization-NMF) μπορεί να θεωρηθεί ως ειδική περιπτωση της NTF, όταν $k = 2$. Οι παράγοντες που προκύπτουν απο τις μεθόδους που προαναφέρθηκαν εχουν συχνά φυσική ή ψυχοφυσιολογική ερμηνεία [32].

Παραγοντοποίηση μη αρνητικών πινάκων -NMF

Οι αλγόριθμοι παραγοντοποίησης μη αρνητικών πινάκων (non-negative matrix factorization - NMF) βρίσκουν τη μη αρνητική παραγοντοποίηση ενός δεδομένου μη αρνητικού πίνακα. Δοθέντος μη αρνητικού πίνακα $\mathbf{V}$ διαστάσεων $n \times m$, οι NMF αλγόριθμοι δίνουν ως έξοδο τους μη αρνητικούς πίνακες $\mathbf{W}$ και $\mathbf{H}$ έτσι ώστε:

$$\mathbf{V} \approx \mathbf{WH} \tag{5.10}$$

όπου ο πίνακας $\mathbf{W}$ έχει διαστάσεις $n \times r$, και ο πίνακας $\mathbf{H}$ έχει διαστάσεις $r \times m$, με τον περιορισμό $(n + m)r < nm$, έτσι ώστε οι πίνακες $\mathbf{W}$ και $\mathbf{H}$ να είναι μικρότερης διάστασης από τον αρχικό πίνακα $\mathbf{V}$, που έχει ως αποτέλεσμα μια συμπιεσμένη εκδοχή του αρχικού πίνακα δεδομένων.

Η (5.10) μπορεί να ξαναγραφτεί ανά στήλες, έτσι ώστε $\mathbf{v}_i \approx \mathbf{Wh}_i$, $i = 1, 2, \ldots, m$, όπου $\mathbf{v}_i$ και $\mathbf{h}_i$ είναι οι αντίστοιχες στήλες των $\mathbf{V}$ και $\mathbf{H}$. Με άλλα λόγια, κάθε διάνυσμα πληροφορίας $\mathbf{v}_i$ αναπαρίσταται σαν γραμμικός συνδυασμός των στηλών του $\mathbf{W}$, ζυγισμένο με τα στοιχεία του $\mathbf{H}$. Αν θεωρήσουμε το r σαν αριθμό κλάσεων, το στοιχείο w_{ij} αναπαριστά τον βαθμό με τον οποίον το i-οστό χαρακτηριστικό ανήκει στην j-οστή κλάση δεδομένων, $j = 1, 2, \ldots, r$, ενώ το στοιχείο h_{jk} δείχνει τον βαθμό με τον οποίο το διάνυσμα πληροφορίας $\mathbf{v}_i$ ανήκει στην κλάση j.

Οι πρώτοι αλγόριθμοι παραγοντοποίησης θετικών πινάκων (positive matrix factorization) προτάθηκαν το 1994, και οι πρώτοι αλγόριθμοι παραγοντοποίησης μη αρνητικών πινάκων το 1999. Για την παραγοντοποίηση NMF χρησιμοποιούνται αλγόριθμοι που βασίζονται σε επαναληπτικές αντικαταστάσεις (update rules) στους πίνακες $\mathbf{W}$ και $\mathbf{H}$. Για να βρεθεί μια προσεγγιστική παραγοντοποίηση του πίνακα $\mathbf{V}$ πρέπει να οριστούν συναρτήσεις κόστους που ορίζουν την ποιότητα της προσέγγισης. Μία τέτοια συνάρτηση μπορεί να κατασκευαστεί χρησιμοποιώντας αποστάσεις μεταξύ δύο μη αρνητικών πινάκων $\mathbf{A} = [a_{ij}]$ και $\mathbf{B} = [b_{ij}]$. Ένα χρήσιμο μέτρο είναι το τετράγωνο της Ευκλείδειας απόστασης μεταξύ των $\mathbf{A}$ και $\mathbf{B}$:

$$||\mathbf{A} - \mathbf{B}||^2 = \sum_{ij} (a_{ij} - b_{ij})^2. \tag{5.11}$$

η οποία έχει σαν κάτω όριο το μηδέν και εμφανώς μηδενίζεται όταν $\mathbf{A} = \mathbf{B}$. Άλλο ένα χρήσιμο μέτρο είναι η Kullback-Leibler απόκλιση (KL divergence) των πινάκων:

$$D_{KL}(\mathbf{A}||\mathbf{B}) = \sum_{ij}\left(a_{ij}\log\frac{a_{ij}}{b_{ij}} - a_{ij} + b_{ij}\right). \tag{5.12}$$

Όπως και η Ευκλείδεια απόσταση, έχει σαν κάτω όριο το μηδέν, και μηδενίζεται όταν $\mathbf{A} = \mathbf{B}$. Δεν είναι θεωρείται απόσταση, γιατί δεν είναι συμμετρική, και αναφέρεται ως απόκλιση του $\mathbf{A}$ από το $\mathbf{B}$. Άρα θεωρούμε δύο διαφορετικές υλοποιήσεις της μεθόδου NMF σαν προβλήματα βελτιστοποίησης: Η πρώτη είναι η ελαχιστοποίηση του $||\mathbf{V} - \mathbf{WH}||^2$ σε σχέση με τα $\mathbf{W}$ και $\mathbf{H}$, με τον περιορισμό $\mathbf{W}, \mathbf{H} \geqslant 0$. Το δεύτερο πρόβλημα είναι η ελαχιστοποίηση του $D(\mathbf{V}||\mathbf{WH})$ σε σχέση με τα $\mathbf{W}$ και $\mathbf{H}$, με τον περιορισμό $\mathbf{W}, \mathbf{H} \geqslant 0$.

Αποδεικνύεται ότι οι παρακάτω επαναληπτικοί πολλαπλασιαστικοί κανόνες συνδυάζουν ευκολία υλοποίησης και ταχύτητα για την λύση των παραπάνω προβλημάτων [74]:

Πρόβλημα 1: Η Ευκλείδεια απόσταση $||\mathbf{V} - \mathbf{WH}||^2$ είναι μη αυξανόμενη στους επαναληπτικούς κανόνες:

$$h_{\alpha j} \leftarrow h_{\alpha j}\frac{(\mathbf{W}^T\mathbf{V})_{\alpha j}}{(\mathbf{W}^T\mathbf{WH})_{\alpha j}} \qquad w_{i\alpha} \leftarrow w_{i\alpha}\frac{(\mathbf{VH}^T)_{i\alpha}}{(\mathbf{WHH}^T)_{i\alpha}}. \tag{5.13}$$

Πρόβλημα 2: Η απόκλιση Kullback-Leibler $D_{KL}(\mathbf{V}||\mathbf{WH})$ είναι μη αυξανόμενη στους επαναληπτικούς κανόνες:

$$h_{\alpha j} \leftarrow h_{\alpha j}\frac{\sum_i \frac{w_{i\alpha}v_{ij}}{(\mathbf{WH})_{ij}}}{\sum_k w_{k\alpha}} \qquad w_{i\alpha} \leftarrow w_{i\alpha}\frac{\sum_j \frac{h_{\alpha j}v_{ij}}{(\mathbf{WH})_{ij}}}{\sum_j h_{\alpha j}}. \tag{5.14}$$

όπου $\alpha = 1, 2...r,\ \ i = 1, 2...n,\ \ j = 1, 2...m$. Όσον αφορά την εφαρμογή του βασικού NMF, οι πίνακες $\mathbf{W}$ και $\mathbf{H}$ μπορούν να αρχικοποιηθούν με τυχαίες μη αρνητικές τιμές, και ο Lee αρχικά ισχυρίζεται στο [74] ότι ο NMF θα συγκλίνει σε τοπικό ελάχιστο. Ο Chu στο [29] καταρρίπτει τον ισχυρισμό του Lee και αποδεικνύει ότι ο NMF δεν συγκλίνει διότι δε πληρούνται οι συνθήκες Karush-Kuhn-Tucker - KKT [13]. Ο Lin προτείνει στο [77] αλγορίθμους παραγοντοποιήσης μη αρνητικών τανυστών χρησιμοποιώντας την μέθοδο βελτιστοποίησης της *προβαλλομένης κλίσης* (projected gradient). Ο Lin εγγυάται ότι ο αλγόριθμός του συγκλίνει σε στασιμο σημείο (stationary point).

Επεκτάσεις του βασικού NMF αποτελούν η επονομαζόμενη τοπική παραγοντοποίηση μη αρνητικών πινάκων (local non-negative matrix factorization - LNMF), για εκμάθηση χωρικά τοπικών αναπαραστάσεων προτύπων [81] καθώς και ο αραιός NMF (sparse non-negative matrix factorization - SNMF) [61]. Βασικό σκεπτικό πίσω από τον αραιό NMF είναι ότι ενώ η NMF μέθοδος είναι επιτυχής στην παραγοντοποίηση πινάκων, δεν θέτει όμως περιορισμούς πυκνότητας των δεδομένων στους πίνακες. Σαν αποτέλεσμα, δεν είναι σε θέση να πραγματοποιήσει παραγοντοποίηση σε έναν πίνακα $\mathbf{V}$ που έχει τοπικά αραιά χαρακτηριστικά στα δεδομένα του.

Αν και η NMF έχει χρησιμοποιηθεί με επιτυχία σε μια πληθώρα προβλημάτων όπως η ανάλυση εικόνας, η ομαδοποίηση εγγράφων, η αναγνώριση μουσικών οργάνων [10], ως μέθοδος μείωσης διαστάσεων και αλλού, έχει δυο βασικά μειονεκτήματα. Το πρώτο είναι ότι όταν εφαρμόζεται σε δεδομένα που φυσικά αναπαρίστανται από πίνακες (π.χ. εικόνες) ή από τανυστές τρίτης ή και μεγαλύτερης τάξης (π.χ. βίντεο) λόγω της μετατροπής των πολυδιάστατων δεδομένων σε διανύσματα, η εσωτερική τους δομή καταστρέφεται. Το δεύτερο μειονέκτημα της NMF είναι ότι οι μη αρνητικές αποσυνθέσεις που προκύπτουν είναι εν γένει μη μοναδικές [37, 107].

Συνεπώς μια γενίκευση της NMF, τέτοια ώστε να μπορεί να αναπαραστήσει την εσωτερική δομή πολυδιάστατων μη αρνητικών δεδομένων με πιο φυσικό τρόπο οδηγεί στην παραγοντοποίηση μη αρνητικών τανυστών NTF. Αξίζει να σημειωθεί ότι οι παραγοντοποιήσεις τανυστών με τάξη μεγαλύτερη από 2 είναι μοναδικές υπό ήπιες προϋποθέσεις (mild assumptions) [108].

Στόχος της NTF είναι να εκτιμήσει τα $\mathbf{u}_i^j$ γνωρίζοντας μόνο τον τανυστή $\mathcal{D}$. Λύση στο πρόβλημα της NTF μπορεί να βρεθεί με το να οριστεί μια κατάλληλη αντικειμενική συνάρτηση κόστους (objective function) η οποία και να ελαχιστοποιηθεί. Μπορούν να οριστούν αρκετές αντικειμενικές συναρτήσεις κόστους και να χρησιμοποιηθούν αρκετές διαδικασίες ελαχιστοποίησης των συναρτήσεων κόστους, πράγμα που οδηγεί σε διαφορετικούς αλγορίθμους NTF.

Από το 2001 και έπειτα, έγιναν οι πρώτες απόπειρες δημιουργίας αλγορίθμων παραγοντοποίησης μη αρνητικών τανυστών. Οι περισσότεροι αλγόριθμοι NTF αναφέρονται σε εφαρμογές επεξεργασίας εικόνων, όπου χρησιμοποιούνται τανυστές 3 διαστάσεων.

Ο Welling το 2001 πρότεινε μία γενίκευση του αλγορίθμου θετικής παραγοντοποίησης πινάκων για τανυστές N-οστής τάξης [131]. Ο αλγόριθμος ονομάστηκε παραγοντοποίηση θετικών τανυστών (positive tensor factorization). Στο μοντέλο που προτείνεται, θεωρείται ένας τανυστής $\mathrm{D} \in \mathbb{R}_+^{I_1 \times I_2 \times \cdots I_N}$. Το μοντέλο εκφράζει τα στοιχεία του $\mathcal{D}$ ως:

$$\mathcal{D}_{i_1 i_2 \ldots i_N} = \sum_{j=1}^{K} u^j_{1(i_1)} u^j_{2(i_2)} \cdots u^j_{N(i_N)}, \tag{5.15}$$

όπου $u^j_{1(i_1)}$ είναι το i_1-οστό στοιχείο του διανύσματος $\mathbf{u}_1^j$. Ο βασικός σκοπός του αλγορίθμου είναι η ελαχιστοποίηση του λάθους:

$$RE = \sum_{i_1 i_2 \ldots i_N}^{I_1 I_2 \cdots I_N} \left(\mathcal{D}_{i_1 i_2 \ldots i_N} - \sum_{j=1}^{K} u^j_{1(i_1)} u^j_{2(i_2)} \cdots u^j_{N(i_N)} \right)^2. \tag{5.16}$$

Για την ελαχιστοποίηση του λάθους, χρησιμοποιείται ο παρακάτω επαναληπτικός πολλαπλασιαστικός κανόνας:

$$\mathbf{U}_i = \tilde{\mathbf{U}}_i * \frac{\mathbf{D}_{(i)}\mathbf{Z}}{\tilde{\mathbf{U}}_i \mathbf{Z}^T \mathbf{Z}}. \tag{5.17}$$

όπου $\mathbf{Z} = \sum_{j=1}^{K} u^j_{1(i_1)} u^j_{2(i_2)} \dots u^j_{i-1(i_{i-1})} u^j_{i+1(i_{i+1})} \dots u^j_{N(i_N)} = \mathbf{U}_N \odot \dots \odot \mathbf{U}_{i+1} \odot \mathbf{U}_{i-1} \odot \dots \mathbf{U}_1$. Το σύμβολο $\odot$ δηλώνει το γινόμενο Khatri-Rao, ενώ το σύμβολο $*$ το γινόμενο Hadamard. Όπου οι πίνακες $\mathbf{U}_i \in \mathbb{R}_+^{I_i \times k}$, $i = 1, 2, \dots, N, j = 1, 2, \dots k$ έχουν την ακόλουθη δομή $\mathbf{U}_i = [\mathbf{u}_i^1 | \dots | \mathbf{u}_i^k]$, όπου το διάνυσμα $\mathbf{u}_i^j$ είναι η j-οστή στήλη του πίνακα $\mathbf{U}_i$,

Προτείνεται επίσης και μία συνάρτηση κόστους που αποτελεί γενίκευση της συνάρτησης που προτάθηκε από τον Lee [74] (παραλλαγή απόκλισης Kullback-Leibler).

Το 2005 ο Lim προτείνει μία πολυγραμμική παραλλαγή του NMF, η οποία ουσιαστικά χρησιμοποιεί τον αλγόριθμο PARAFAC με επιπλέον περιορισμό μη αρνητικότητας [82]. Αναφέρεται στο πρόβλημα της εύρεσης ελάχιστου k για το οποίο μία τέτοια αποσύνθεση είναι εφικτή. Σύμφωνα με τον Lim, η ιδανική παραγοντοποίηση μη αρνητικού τανυστή δίνεται από το K όπου:

$$\arg\min_{K} \left\{ ||\mathcal{D} - \sum_{j=1}^{K} \mathbf{u}_1^j \otimes \mathbf{u}_2^j \otimes \dots \otimes \mathbf{u}_N^j||_F, \ \mathbf{u}_i^j \in \mathbb{R}_+^{I_i} \right\}. \tag{5.18}$$

Όμως, δεν είναι πάντα εφικτή η εύρεση του ιδανικού K, άρα αναζητείται η μη ιδανική λύση στην οποία, δοθέντος K, προσεγγίζει έναν τανυστή ως άθροισμα εξωτερικών γινομένων διανυσμάτων.

Οι Shashua και Hazan πρότειναν το 2005 μία γενίκευση του αλγορίθμου NMF για τανυστές N-οστής τάξης [117]. Η μέθοδος ονομάστηκε παραγοντοποίηση μη αρνητικών τανυστών (non-negative tensor factorization - NTF). Το προτεινόμενο μοντέλο βασίζεται στην ιδέα της αποσύνθεσης τανυστή σε άθροισμα εξωτερικού γινομένου διανυσμάτων παραπέμποντας σε μοντέλα ορθογώνιας αποσύνθεσης τανυστών [66] ή σε μοντέλα εύρεσης βαθμού τανυστή (βλ. Ενότητα 3.3.9). Το μοντέλο είναι:

$$\mathcal{D} = \sum_{j=1}^{K} \mathbf{u}_1^j \otimes \mathbf{u}_2^j \otimes \dots \otimes \mathbf{u}_N^j, \tag{5.19}$$

όπου $\mathbf{u}_i^j \in \mathbb{R}^{I_i}$. Το πρόβλημα που επιχειρείται να λυθεί είναι το:

$$\min_{\mathbf{u}_i^j} \frac{1}{2} ||\mathcal{D} - \sum_{j=1}^{K} \mathbf{u}_1^j \otimes \mathbf{u}_2^j \otimes \dots \otimes \mathbf{u}_N^j||_F^2, \quad \mathbf{u}_i^j \geqslant 0, \ i = 1, \dots, N, \ j = 1, \dots, K \tag{5.20}$$

όπου $||.||_F^2$ είναι η νόρμα Frobenius. Η εύρεση του πολλαπλασιαστικού επαναληπτικού κανόνα βρίσκεται λύνοντας την εξίσωση $\frac{\partial f}{\partial u^j_{i(i_i)}} = 0$, όπου f είναι οι όροι στην (5.20). Ο επαναληπτικός κανόνας που προκύπτει:

$$u^p_{i(l)} \leftarrow u^p_{i(l)} \frac{\sum_{i_1 \dots i_{i-1} i_{i+1} \dots i_N}^{I_1 \dots I_{i-1} I_{i+1} \dots I_N} \mathcal{D}_{i_1 \dots l \dots i_N} \prod_{m \neq i} u^p_{m(i_m)}}{\sum_{j=1}^{K} u^j_{i(l)} \prod_{m \neq i} (\mathbf{u}_m^{jT} \mathbf{u}_m^p)}. \tag{5.21}$$

Το πρόβλημα που επιχειρεί να λύσει ο Shashua είναι όμοιο με το πρόβλημα του Welling [131], με τη μόνη διαφορά στον περιορισμό (μη αρνητικές τιμές αντί θετικών τιμών). Ο επαναληπτικός κανόνας που προκύπτει είναι παραπλήσιος με αυτόν της (5.17).

Οι Hazan και Shashua συνέχισαν την έρευνα σχετικά με τις NTF μεθόδους, αυτή τη φορά περιορίζοντας τον αριθμό των διαστάσεων στις 3 [51]. Προτάθηκε ένας αλγόριθμος για NTF χρησιμοποιώντας την 'σχετική εντροπία' σαν μέτρο απόστασης, κοινώς την απόκλιση Kullback-Leibler. Ο επαναληπτικός πολλαπλασιαστικός κανόνας που προκύπτει για έναν τανυστή 3 διαστάσεων $\mathcal{D} \in \mathbb{R}_{+}^{I_1 \times I_2 \times I_3}$ είναι:

$$u_{1(l)}^{p} \leftarrow u_{1(l)}^{p} \cdot \frac{\sum_{i_2 i_3}^{I_2 I_3} \mathcal{D}_{1 i_2 i_3} \frac{u_{2(i_2)}^{p} u_{3(i_3)}^{p}}{\sum_{m=1}^{K} u_{1(i_1)}^{m} u_{2(i_2)}^{m} u_{3(i_3)}^{m}}}{\sum_{i_2 i_3}^{I_2 I_3} u_{2(i_2)}^{p} u_{3(i_3)}^{p}}. \tag{5.22}$$

Ο κανόνας δίνεται αντίστοιχα για τα $u_{2(l)}^{p}$ και $u_{3(l)}^{p}$. Οι αλγόριθμοι στο [51] συγκρίνονται με τους NMF και PCA για κωδικοποίηση εικόνων.

Ο Fitzgerald προτείνει έναν αλγόριθμο παραγοντοποιήσης μη αρνητικών τανυστών χρησιμοποιώντας την γενικευμένη απόσταση Kullback-Leibler [43]. Ο επαναληπτικός πολλαπλασιαστικός κανόνας που προκύπτει για έναν τανυστή 3 διαστάσεων $\mathcal{V} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ είναι όμοιος με αυτόν της (5.23) των Hazan και Shashua και ακολουθεί:

$$u_{1(l)}^{p} \leftarrow u_{1(l)}^{p} \cdot \frac{\sum_{i_2 i_3}^{I_2 I_3} \mathcal{D}_{1 i_2 i_3} \frac{u_{2(i_2)}^{p} u_{3(i_3)}^{p}}{\sum_{m=1}^{K} u_{1(i_1)}^{m} u_{2(i_2)}^{m} u_{3(i_3)}^{m}}}{\sum_{i_2 i_3}^{I_2 I_3} u_{2(i_2)}^{p} u_{3(i_3)}^{p}}. \tag{5.23}$$

Να σημειωθεί οτι στο [43] δεν αναφερονται αποδείξεις περι συγκλισης του αλγορίθμου.

Το 2006 οι Heiler και Schnörr πρότειναν μία γενίκευση του αλγορίθμου που είχε προτείνει ο Hoyer [59] για NMF με περιορισμούς αραιότητας [53]. Ο αλγόριθμος προτάθηκε για τανυστές 3 διαστάσεων (με πιθανή εφαρμογή την κωδικοποίηση εικόνων). Το πρόβλημα που τίθεται είναι όμοιο με αυτό που προτάθηκε από τους Welling [131] και Hazan [117, 51]:

$$\min_{\mathbf{u}_i^j} \frac{1}{2} ||\mathcal{D} - \sum_{j=1}^{K} \mathbf{u}_1^j \otimes \mathbf{u}_2^j \otimes \mathbf{u}_3^j||_F^2, \qquad \mathbf{u}_i^j \geqslant 0 \tag{5.24}$$

με τη διαφορά ότι τίθεται ένας επιπλέον περιορισμός αραιότητας. Ο Heiler ορίζει την αραιότητα όπως την ορίζει ο Albright στο [1]. Ο επιπλέον περιορισμός είναι:

$$s_i^{min} \leqslant \mathrm{spar}(u_i) \leqslant s_i^{max}. \tag{5.25}$$

Οι παράμετροι s_i^{min} και s_i^{max} είναι πραγματικοί αριθμοί στο διάστημα $[0, 1]$ και δίνονται από τον χρήστη, ανάλογα με την εφαρμογή. Ο αλγόριθμος ονομάστηκε SMA (sparsity maximization algorithm) και υπάρχει σε μορφή ψευδοκώδικα στο [53]. Αποδεικνύεται ότι ο αλγόριθμος συγκλίνει σε πεπερασμένο χρονικό διάστημα σε τοπικό ελάχιστο. Στο [53] πραγματοποιήθηκαν πειράματα εφαρμογής του SMA σε αναγνώριση προσώπων.

Το 2006 ο Μπουτσίδης πρότεινε έναν αλγόριθμο για την αποσύνθεση ενός μη αρνητικού τανυστή [17]. Ο αλγόριθμος ονομάστηκε PALSIR (projected alternating least squares with initialization and regularization). Το πρόβλημα που θέτει είναι το:

$$\min_{\mathbf{u}_i^j} \frac{1}{2}||\mathcal{D} - \sum_{j=1}^{K} \mathbf{u}_1^j \otimes \mathbf{u}_2^j \otimes \mathbf{u}_3^j||_F^2, \qquad \mathbf{u}_i^j \geqslant 0. \tag{5.26}$$

Για την επίλυση του προβλήματος, χρησιμοποιείται ο αλγόριθμος ALS. Από τα διανύσματα $\mathbf{u}_i^j$ δημιουργούνται οι πίνακες $\mathbf{U}_i \in \mathbb{R}_+^{I_i \times K}$. Σε κάθε επανάληψη, για κάθε $i = 1, 2, 3$, λύνεται το σύστημα της (5.26) με τη μέθοδο των ελαχίστων τετραγώνων. Επίσης προτείνεται η ιδέα εφαρμογής μεθόδων αρχικοποίησης των τιμών για μεγαλύτερη ταχύτητα και καλύτερη σύγκλιση. Προτείνεται η χρήση αρχικοποιήσεων που υπάρχουν για την μέθοδο PARAFAC, καθώς και πολυγραμμικές επεκτάσεις των μεθόδων αρχικοποίησης του NMF που πρότεινε ο Albright [1]. Παρόλα αυτά, δεν γίνεται κάποια προσπάθεια μελέτης των μεθόδων αρχικοποίησης. Να σημειωθεί ότι στο [17] δεν δίνεται απόδειξη για τον αλγόριθμο και την σύγκλισή του.

Το 2007 ο Cichocki πρότεινε έναν αλγόριθμο που πραγματοποιεί παραγοντοποίηση μη αρνητικών τανυστών με περιορισμούς αραιότητας [27]. Ο αλγόριθμος προτείνεται μόνο για τανυστές 3 διαστάσεων. Το προτεινόμενο μοντέλο είναι το παρακάτω:

$$\mathbf{D}_k = \mathbf{A}\mathbf{V}_k\mathbf{S}_k + \mathbf{E}_k, \qquad k = 1, 2, \ldots, I_3. \tag{5.27}$$

Τα $\mathbf{D}_k = \mathbf{D}_{:,:,k} \in \mathbb{R}^{I_1 \times I_2}$ είναι τομές (frontal slices) του τανυστή $\mathcal{D} \in \mathbb{R}_+^{I_1 \times I_2 \times I_3}$ (θεωρούμε I_3 τομές). Ο $\mathbf{A} \in \mathbb{R}^{I_1 \times R}$ είναι ο πίνακας βάσης (ή μίξης) που αναπαριστά τους παράγοντες που συνθέτουν τα δεδομένα. Ο $\mathbf{V}_k \in \mathbb{R}^{R \times R}$ είναι διαγώνιος πίνακας που έχει στην κύρια διαγώνιο του στοιχεία του πίνακα $\mathbf{D} \in \mathbb{R}^{I_3 \times R}$. Οι πίνακες $\mathbf{S}_k \in \mathbb{R}^{R \times I_2}$ αναπαριστούν τις πηγές (ή κρυφές συνιστώσες) των δεδομένων. Τέλος, οι πίνακες $\mathbf{E}_k \in \mathbb{R}^{I_1 \times I_2}$ είναι οι τομές του τανυστή $\mathcal{E} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ που περιέχει τα λάθη ή τον θόρυβο, ανάλογα με την εφαρμογή.

Ο Cichocki στο [31] χρησιμοποιεί σαν συναρτήσεις κόστους τις α- και β-αποκλίσεις. Η α-απόκλιση ορίζεται ως:

$$D_k^{(\alpha)}(\mathbf{D}_k||\mathbf{A}\mathbf{S}_k) = \frac{1}{\alpha(\alpha-1)} \sum_{itk} (v_{itk}^{\alpha}[\mathbf{A}\mathbf{S}_k]_{it}^{1-\alpha} - \alpha v_{itk} + (\alpha-1)[\mathbf{A}\mathbf{S}_k]_{it}). \tag{5.28}$$

Να σημειωθεί ότι για $\alpha = 2$, η (5.28) γίνεται η απόσταση του Pearson. Για $\alpha = 0.5$, η (5.28) γίνεται η απόσταση του Hellinger. Για $\alpha = -1$, η (5.28) γίνεται η απόσταση του Neyman. Για $\alpha \to 1$ γίνεται η γενικευμένη απόκλιση Kullback-Leibler (ονομάζεται και I-απόκλιση). Χρησιμοποιώντας μία μετασχηματισμένη εκδοχή της ελάττωσης κατά κλίση (gradient descent), οι επαναληπτικοί κανόνες που προκύπτουν είναι:

$$s_{rtk} \leftarrow s_{rtk} \cdot \left(\frac{\sum_{i=1}^{I_1} a_{ir}(v_{itk}/[\mathbf{A}\mathbf{S}_k]_{it})^{\alpha}}{\sum_{q=1}^{I_1} a_{qr}} \right)^{1/\alpha}. \tag{5.29}$$

$$a_{ir} \leftarrow a_{ir} \cdot \left(\frac{\sum_{p=1}^{I_2 I_3} (v_{ip}/[\mathbf{AS}_k]_{ip})^{\alpha} s_{rp}}{\sum_{q=1}^{I_2 I_3} s_{rq}} \right)^{1/\alpha}. \tag{5.30}$$

Οι περιορισμοί αραιότητας μπορούν να επιτευχθούν με κατάλληλο μη γραμμικό μετασχηματισμό της μορφής $s_{rtk} \leftarrow (s_{rtk})^{1+\gamma}$, όπου γ συντελεστής αραιότητας.

Η β-απόκλιση μπορεί να θεωρηθεί σαν συμπληρωματική απόκλιση με την α. Η συνάρτηση κόστους του NTF αλγορίθμου που προτείνει ο Cichocki χρησιμοποιώντας β-απόκλιση είναι:

$$D_k^{(\beta)}(\mathbf{V}_k || \mathbf{AS}_k) = \sum_{it} \left(v_{itk} \frac{v_{itk}^{\beta} - [\mathbf{AS}_k]_{it}^{\beta}}{\beta(\beta+1)} + [\mathbf{AS}_k]_{it}^{\beta} \frac{[\mathbf{AS}_k]_{it}^{\beta} - v_{itk}}{\beta+1} \right) + \alpha_{\mathbf{S}_k} ||\mathbf{S}_k||_{L1}, \tag{5.31}$$

όπου α_A είναι παράμετροι κανονικοποίησης που ελέγχουν τον βαθμό αραιότητας των πινάκων $\mathbf{S}$ και $\mathbf{A}$. Στην περίπτωση που $\beta = 1$ και $\alpha_{\mathbf{S}_k} = 0$, η (5.31) εκφράζει την νόρμα Frobenius. Όταν $\beta \to 0$ εκφράζεται η γενικευμένη απόκλιση Kullback-Leibler. Τέλος, όταν $\beta \to -1$, η (5.31) μετατρέπεται στην απόσταση Itakura-Saito. Οι πολλαπλασιαστικοί κανόνες που προκύπτουν χρησιμοποιώντας τις β-αποκλίσεις είναι:

$$s_{rtk} \leftarrow s_{rtk} \cdot \frac{[\sum_{i=1}^{I_1} a_{ir} (v_{itk}/[\mathbf{AS}_k]_{it}^{1-\beta}) - \alpha_{\mathbf{S}_k}]_{\epsilon}}{\sum_{p=1}^{I_1} a_{ir} [\mathbf{AS}_k]_{it}^{\beta}} \tag{5.32}$$

$$a_{ir} \leftarrow a_{ir} \cdot \frac{[\sum_{p=1}^{I_2 I_3} (v_{ip}/[\mathbf{AS}_k]_{it}^{1-\beta}) s_{rp} - \alpha_A]_{\epsilon}}{\sum_{p=1}^{I_2 I_3} [\mathbf{AS}_k]_{it}^{\beta} s_{rp}} \tag{5.33}$$

όπου $[x]_{\epsilon} = \max\{\epsilon, x\}$ για την αποφυγή μηδενικών τιμών.

Ο Cichocki επεκτείνει την έρευνα στους NTF αλγορίθμους στο [30], όπου προτείνει εναλλακτικούς αλγορίθμους για τις ίδιες συναρτήσεις κόστους. Προτείνει αλγορίθμους βασισμένους στην ALS [32] τεχνικές και σε τεχνικές βελτισποποιήσης όπως Alternating Interior-Point Gradient και Quasi-Netwon. Να σημειωθεί ότι οι αλγόριθμοι που δίνονται αφορούν μόνο δεδομένα 3 διαστάσεων. Δεν δίνονται αποδείξεις σχετικά με την εύρεση των πολλαπλασιαστικών κανόνων από τις συναρτήσεις κόστους. Αξίζει να σημειωθει ότι το μοντέλο (5.27) αφενός δεν μπορεί να γενικευτεί για περισσότερες διαστάσεις, αφετέρου δεν μπορεί να γίνει αναγωγή του μοντέλου στην περίπτωση 2 διαστάσεων, όπου θεωρητικά θα κατέληγε στο μοντέλο του NMF.

Ο Μπενέτος στο [8, 9] προτείνει μια κλάση αλγορίθμων χρησιμοποιώντας τις αποκλίσεις Bregman οι οποίες αποτελούν μία οικογένεια συναρτήσεων που εκφράζουν διαφορετικά μέτρα [18]. Η ελαχιστοποίηση των απκλίσεων προκύπτει χρησιμοποιόντας βοηθητικές συναρτήσεις.Οι πολλαπλασιαστικοί κανόνες ενημέρωσης που προκύπτουν για κάθε απόκλιση Bregman είναι:

Ο επαναληπτικός κανόνας ενημέρωσης για παραγοντοποίηση μη αρνητικών τανυστών χρησιμοποιώντας την απόκλιση Kullback - Leibler ορίζεται από την (5.34).

$$u^p_{i(l)} \leftarrow \tilde{u}^p_{i(l)} \cdot \exp\left(\frac{\sum(u^p_{1(i_1)} \dots u^p_{i-1(i_{i-1})} u^p_{i+1(i_{i+1})} \dots u^p_{N(i_N)}) \cdot \log(\frac{\mathcal{D}_{i_1 \dots l \dots i_N}}{\sum_{m=1}^{k} u^m_{1(i_1)} \dots \tilde{u}^m_{i(l)} \dots u^m_{N(i_N)}})}{\sum u^p_{1(i_1)} \dots u^p_{i-1(i_{i-1})} u^p_{i+1(i_{i+1})} \dots u^p_{N(i_N)}}\right). \tag{5.34}$$

Ο επαναληπτικός κανόνας για NTF χρησιμοποιώντας την νόρμα Frobenious ορίζεται από την (5.35).

$$u^p_{i(l)} \leftarrow \tilde{u}^p_{i(l)} \cdot \frac{\sum(u^p_{1(i_1)} \dots u^p_{i-1(i_{i-1})} u^p_{i+1(i_{i+1})} \dots u^p_{N(i_N)}) \cdot \mathcal{D}_{i_1 \dots l \dots i_N}}{\sum(u^p_{1(i_1)} \dots u^p_{i-1(i_{i-1})} u^p_{i+1(i_{i+1})} \dots u^p_{N(i_N)}) \cdot (\sum_{m=1}^{k} u^m_{1(i_1)} \dots \tilde{u}^m_{i(l)} \dots u^m_{N(i_N)})}. \tag{5.35}$$

Ο Μπενέτος τέλος, προτείνει τον πολλαπλασιαστικό κανόνα ενημέρωσης χρησιμοποιώντας την απόσταση Itakura - Saito οπου ορίζεται ως:

$$u^p_{i(l)} \leftarrow \tilde{u}^p_{i(l)} \cdot \frac{\sum_{i_1,\dots,i_{i-1},i_{i+1},\dots,i_N}^{I_1 \cdots I_{i-1} I_{i+1} \cdots I_N} \frac{u^p_{1(i_1)} \dots u^p_{i-1(i_{i-1})} u^p_{i+1(i_{i+1})} \dots u^p_{N(i_N)}}{\sum_{m=1}^{k} u^m_{1(i_1)} \dots \tilde{u}^m_{i(l)} \dots u^m_{N(i_N)}}}{\sum_{i_1,\dots,i_{i-1},i_{i+1},\dots,i_N}^{I_1 \cdots I_{i-1} I_{i+1} \cdots I_N} \frac{u^p_{1(i_1)} \dots u^p_{i-1(i_{i-1})} u^p_{i+1(i_{i+1})} \dots u^p_{N(i_N)}}{\mathcal{D}_{i_1 \dots l \dots i_N}}}. \tag{5.36}$$

Στην παρούσα διατριβή θα χρησιμοποιήσουμε τον αλγόριθμο NTF με την νόρμα Frobenious όπου ορίζεται στην (5.35) [8] ως πολυγραμμική μέθοδο εξαγωγής χαρακτηριστικών. Κοινό χαρακτηριστικό των παραπάνω αλγορίθμων είναι ότι αποδεικνύεται μεν ότι οι η συνάρτηση κόστους δεν αυξάνεται σε κάθε επανάληψη, αλλά όχι ότι οι NTF αλγόριθμοι συγκλίνουν σε τοπικό ελάχιστο.

5.3.5 Παραγοντοποίηση Μη-Αρνητικών Τανυστών Χρησιμοποιώντας Προβαλλόμενα Διανύσματα Κλίσης

Για την εύρεση λύσης στο πρόβλημα της παραγοντοποίησης μη αρνητικού τανυστή N-τάξης, όπως αυτό ορίζεται στην (5.9), μια συνηθισμένη προσέγγιση είναι η ελαχιστοποίηση του κόστους Ελάχιστων Τετραγώνων (Least Square Error - LSE), το οποίο ορίζεται ως εξής:

$$\varphi(\mathcal{D}, \sum_{j=1}^{k} \mathbf{u}^j_1 \otimes \mathbf{u}^j_2 \otimes \cdots \otimes \mathbf{u}^j_N) = \frac{1}{2} \|\mathcal{D} - \sum_{j=1}^{k} \mathbf{u}^j_1 \otimes \mathbf{u}^j_2 \otimes \cdots \otimes \mathbf{u}^j_N\|^2_F. \tag{5.37}$$

Συνεπώς η λύση στο πρόβλημα της παραγοντοποιήσης μη αρνητικού τανυστή N-τάξης (5.9) προκύπτει από την επίλυση του παρακάτω πρόβληματος ελαχιστοποίησης:

$$\min_{\mathbf{u}^j_i \geqslant 0} \varphi(\mathcal{D}, \sum_{j=1}^{k} \mathbf{u}^j_1 \otimes \mathbf{u}^j_2 \otimes \cdots \otimes \mathbf{u}^j_N) \; s.\,t.\; \mathbf{u}^j_i \geqslant 0. \tag{5.38}$$

Ορίζοντας N πίνακες $\mathbf{U}_i \in \mathbb{R}_+^{I_i \times k}$, $i = 1, 2, \ldots, N, j = 1, 2, \ldots k$ με την ακόλουθη δομή $\mathbf{U}_i = [\mathbf{u}_i^1 | \ldots | \mathbf{u}_i^k]$, όπου το διάνυσμα $\mathbf{u}_i^j$ είναι η j-οστή στήλη του πίνακα $\mathbf{U}_i$, η αντικειμενική συνάρτηση κόστους (5.37) ισοδυναμέι με N συναρτήσεις κόστους, έστω:

$$\varphi_i(\mathbf{D}_{(i)}, \mathbf{U}_i\mathbf{Z}^T) = \frac{1}{2}\|\mathbf{D}_{(i)} - \mathbf{U}_i\mathbf{Z}^T\|_F^2, i = 1, 2, \ldots, N. \tag{5.39}$$

$$\mathbf{Z} = \mathbf{U}_N \odot \ldots \odot \mathbf{U}_{i+1} \odot \mathbf{U}_{i-1} \odot \ldots \mathbf{U}_1. \tag{5.40}$$

$$\tag{5.41}$$

Οι συναρτήσεις κόστους που ορίζονται στην (5.39) δεν είναι κυρτές λόγω της παρουσίας του γινομένου $\mathbf{U}_i\mathbf{Z}^T$. Ως εκ τούτου, είναι δύσκολο να βρούμε ολικό ελάχιστο. Παρά ταύτα, κάθε συνάρτηση κόστους που ορίζεται στην (5.39) είναι ξεχωριστά κυρτή για κάθε ομάδα μεταβλητων $\mathbf{U}_i, i = 1, 2, \ldots, N$. Το γεγονός αυτό καθιστά δυνατό να χρησιμοποιήσουμε το σχήμα των Εναλλασσόμενων Ελάχιστων Τετραγώνων (Alternating Least Squares - ALS) προκειμένου να ελαχιστοποιήσουμε την (5.37).

Η ιδέα πίσω από τα ALS είναι απλή: ενημερώνονται επαναληπτικά οι τιμές μιας ομάδας (block) μεταβλητών, υπό τους συγκεκριμένους περιορισμούς, ενώ οι υπόλοιπές ομάδες μεταβλητών παραμένουν σταθερές. Η διαδικασία επαναλαμβάνεται μέχρις ότου επιτευχθεί σύγκλιση σε στάσιμο σημείο. Το στάσιμο σημείο δεν είναι απαραίτητα τοπικό ελάχιστο, αλλά ένα τοπικό ελάχιστο είναι πάντα στάσιμο. Η σύγκλιση σε στάσιμο σημείο εξασφαλίζεται από τους ALS αλγορίθμους [13, 77, 50].

Για το πρόβλημα της μη αρνητικής παραγοντοποιήσης τανυστή N-τάξης, έχουμε N-ομάδες μεταβλητών $\mathbf{U}_i, i = 1, 2, \ldots, N$ και η λύση στη (5.38) δίνεται από την ακολουθιακή επίλυση των N υποπροβλημάτων που ορίζονται στην (5.42).

$$\min_{\mathbf{U}_i \geqslant 0} \varphi_i(\mathbf{D}_{(i)}, \mathbf{U}_i\mathbf{Z}^T) \; s.\,t.\; \mathbf{U}_i \geqslant 0, \forall i. \tag{5.42}$$

Η προαναφερθείσα προσέγγιση είναι γνωστή ως *ομαδική κατά συνιστώσα κατάβαση* ("block coordinate descent") ή ομαδική μη-γραμμική Gauss-Seidel μέθοδος ("block non-linear Gauss-Seidel method") [13, 50].

Όταν κρατάμε σταθερό ένα σύνολο μεταβλητών, κάθε υποπρόβλημα που ορίζεται στην (5.42) αποτελείται από ένα σύνολο προβλημάτων ελαχιστοποίησης μη αρνητικών ελαχίστων τετραγώνων (Non-Negative Least Squares - NNLS):

$$\min_{\mathbf{u}_i^j \geqslant 0} \frac{1}{2}\|\mathbf{d}_i^{jT} - \mathbf{Z}\mathbf{u}_i^{jT}\|_F^2 \; s.\,t.\; \mathbf{u}_i^j \geqslant 0, \forall i. \tag{5.43}$$

όπου $\mathbf{d}_i^j$ είναι j-οστή στήλη του πίνακα $\mathbf{D}_{(i)}$ και $\mathbf{u}_i^j$ είναι η άγνωστη j-οστή στήλη του πίνακα $\mathbf{U}_i$. Συνεπώς η εύρεση μη αρνητικής παραγοντοποιήσης τανυστή η οποία έγκειται στην επίλυση των

υποπροβλημάτων που ορίζονται (5.42) ανάγεται στην επίλυση ενός συνόλου NNLS προβλημάτων όπως ορίζονται στην (5.43) για κάθε υποπρόβλημα.

Υπάρχουν αρκετοί αλγόριθμοι επίλυσης NNLS προβλημάτων. Στο παρόν κεφάλαιο θα υιοθετήσουμε μια κλάση αλγορίθμων επίλυσης NNLS προβλημάτων που χρησιμοποιεί *προβαλλόμενα διανύσματα κλίσης* (Projected Gradients -PG). Βασιζόμενοι σε αυτή την μεθοδολογία στις επόμενες ενότητες του κεφαλαίου θα προτείνουμε δύο νέους αλγορίθμους παραγοντοποιήσης μη αρνητικού τανυστή χρησιμοποιώντας προβαλλόμενα διανύσματα κλίσης.

Σε αντίθεση με τους πολλαπλασιαστικούς κανόνες ενημέρωσης, η κλάση των PG αλγορίθμων έχει προσθετικούς κανόνες ενημέρωσης. Οι PG αλγόριθμοι, γενικά, μπορούν να εκφραστούν από τον ακόλουθο επαναληπτικό κανόνα ενημέρωσης:

$$\mathbf{U}_i^{(t+1)} = P_\Omega[\mathbf{U}_i^{(t)} - n_{U_i}^{(t)} \mathbf{P}_{\mathbf{U}_i}^{(t)}], \tag{5.44}$$

όπου $P_\Omega[\zeta]$ είναι η προβολή του ζ στο δυνατό κυρτό σύνολο (convex feasible set) $\Omega = \{\zeta \in \mathbb{R}_+\}$, $\mathbf{P}_{\mathbf{U}_i}^{(t)}$ είναι η κατεύθυνση κατάβασης (descent direction) για το $\mathbf{U}_i$ και $n_{U_i}^{(t)}$ είναι οι ρυθμοί μάθησης (learning rates).

Στην επόμενη ενότητα παρουσιάζονται δυο νέοι NTF αλγόριθμοι που χρησιμοποιούν προβαλλόμενα διανύσματα κλίσης (Projected Gradients) και γενικά θα αναφέρονται ως PG-NTF. Ο πρώτος αλγόριθμος παραγοντοποίησης μη-αρνητικού τανυστή θα αναφέρεται ως Projected Landweber-NTF ή PL-NTF διότι χρησιμοποιεί την προβαλλόμενη Landweber μέθοδο [77, 63] για την επίλυση των υποπροβλημάτων που ορίζονται στην (5.42). Ο δεύτερος αλγόριθμος παραγοντοποίησης μη-αρνητικού τανυστή θα αναφέρεται ως CW-NTF διότι χρησιμοποιεί την ακολουθιακή κατά συνιστώσα μέθοδο (sequential coordinate-wise) [44] για την επίλυση των μη αρνητικών προβλημάτων ελαχίστων τετραγώνων τα οποία συνθέτουν κάθε υποπρόβλημα στην (5.42).

Αλγόριθμος NTF Χρησιμοποιώντας την Μέθοδο του Landweber

Η προβαλλόμενη Landweber μέθοδος (Projected Landweber) υιοθετείται έτσι ώστε να εξάγουμε έναν αλγόριθμο παραγοντοποίησης μη αρνητικών τανυστών, ο οποίος εφεξής θα αναφέρεται ως PL-NTF.

Η προβαλλόμενη Landweber μέθοδος είναι μια μέθοδος κατάβασης κατά το διάνυσμα κλίσης (gradient descent), οπού επαναληπτικά ανανεώνει τις τιμές μιας ομάδας μεταβλητών, έστω τα στοιχεία του πίνακα $\mathbf{U}_i$, ενώ τα στοιχεία του πίνακα $\mathbf{Z}$ παραμένουν αμετάβλητα. Ο πίνακας $\mathbf{Z}$ ορίζεται όπως στην 5.40. Η παραπάνω διαδικασία εγγυάται ότι το οριακό σημείο (limit point) είναι στάσιμο σημείο (stationary point) [44, 63, 13, 50]. Η παραγοντοποίηση μη αρνητικού τανυστή με την προβαλλόμενη Landweber μέθοδο προκύπτει από την επαναλαμβανώμενη επί-

λυση των υποπροβλήματων που ορίζονται στην (5.42). Η μέθοδος απαιτεί τον υπολογισμό των κλίσεων πρώτης τάξης των N συναρτήσεων στη (5.39) όπου δίνονται από την εξίσωση (5.45).

$$\nabla\varphi_i(\mathbf{U}_i) = (\mathbf{U}_i\mathbf{Z}^T - \mathbf{D}_{(i)})\mathbf{Z} \tag{5.45}$$

Αρχικά οι πίνακες $\mathbf{U}_j$ αρχικοποιούνται είτε τυχαία είτε χρησιμοποιώντας κάποια μέθοδο αρχικοποίησης από το [1], κατά τέτοιο τρόπο ώστε τα στοιχεία τους να είναι μη αρνητικά.

Εν συνεχεία επιλύονται επαναληπτικά τα N υποπροβλήματα που ορίστηκαν στην (5.42). Έστω, λοιπόν, η ελαχιστοποίηση του υποπροβλήματος (5.42) ως προς $\mathbf{U}_i$, κρατώντας τον πίνακα $\mathbf{Z}$ σταθερό. Η ελαχιστοποίηση του είναι μια επαναληπτική διαδικασία η οποία επαναλαμβάνεται μέχρι ο πίνακας $\mathbf{U}_i^{(t)}$ αποτελεί στάσιμο σημείο της (5.42). Σε κάθε επανάληψη ένα κατάλληλο μέγεθος βήματος απαιτείται έτσι ώστε να ενημερωθούν οι τιμές του πίνακα $\mathbf{U}_i^{(t)}$. Όταν βρεθούν οι κατάλληλες τιμές ανανέωσης για τα στοιχεία $\mathbf{U}_i$, ελέγχεται η συνθήκη στασιμότητας και εάν ικανοποιείται ο αλγόριθμος σταματά.

Για ένα πλήθος επαναλήψεων $t = 1, 2, \ldots$ κανόνας ενημέρωσης [77] για τον πίνακα $\mathbf{U}_i$ είναι:

$$\mathbf{U}_i^{(t+1)} = P_\Omega[\mathbf{U}_i^{(t)} - n_{U_i}^{(t)}\nabla\varphi_i(\mathbf{U}_i)] \tag{5.46}$$

όπου $n_{U_i^{(t)}} = \beta^{g_t}$και g_t είναι ο πρώτος μη αρνητικός ακέραιος τέτοιος ώστε:

$$\varphi_i(\mathbf{D}_{(i)}, \mathbf{U}_i^{(t+1)}\mathbf{Z}^T) - \varphi_i(\mathbf{D}_{(i)}, \mathbf{U}_i^{(t)})\mathbf{Z}^T) \leqslant \sigma\langle\nabla\varphi_i(\mathbf{U}_i^{(t)}), \mathbf{U}_i^{(t+1)} - \mathbf{U}_i^{(t+1)}\rangle \tag{5.47}$$

όπου $\langle . \rangle$ το εσωτερικό γινόμενο των πινάκων.

Η συνθήκη (5.47) εξασφαλίζει ότι σε κάθε επανάληψη, η τιμή της συνάρτηση $\varphi_i(.)$ θα έχει μειωθεί επαρκώς. Ο κανόνας προβολής $P_\Omega[.] = \max[., 0]$ αναφέρεται στα στοιχεία του πίνακα και εξασφαλίζει ότι αυτά είναι μη αρνητικά σε κάθε ενημέρωση.

Η αναζήτηση κατάλληλης τιμής για $n_{U_i}^{(t)}$ είναι η πιο χρονοβόρα διαδικασία του αλγορίθμου, συνεπώς είναι επιθυμητό να γίνουν οι λιγότερες δυνατές επαναλήψεις. Διαφορές διαδικασίες έχουν προταθεί για την επιλογή και ενημέρωση των τιμών του $n_{U_i^{(t)}}$ [13]. Στην παρούσα διατριβή χρησιμοποιούμε τον Αλγόριθμο 4 από το [77], ενώ οι τιμές β, σ επιλέχθηκαν έτσι ώστε να είναι ίσες με 0.1 και 0.001 ($0 < \beta < 1, 0 < \sigma < 1$), αντίστοιχα. Η αναζήτηση $n_{U_{(i)}}^{(t)}$ επαναλαμβάνεται μέχρις ότου το σημείο $\mathbf{U}^{(j)(t)}$ γίνει στάσιμο.

Μία κοινή συνθήκη ελέγχου στασιμότητας που έχει χρησιμοποιηθεί στην βιβλιογραφία είναι η εξής [77, 67]:

$$\|\nabla^P\varphi_i(\mathbf{U}_i{}^{(t)})\|_F \leqslant \epsilon_{\mathbf{U}_i}\|\nabla\varphi_i(\mathbf{U}_i{}^{(1)})\|_F \tag{5.48}$$

όπου $\nabla^P\varphi_i(\mathbf{U}_i)$ είναι η προβαλλόμενη κλήση και ορίζεται ως:

$$[\nabla^P \varphi_i(\mathbf{U}_i)]_{ij} = \{ \begin{array}{l} [\nabla^P \varphi_i(\mathbf{U}_i)]_{ij}, \text{αν} [\mathbf{U}_i]_{ij} > 0 \\ \min(0, [\nabla^P \varphi_i(\mathbf{U}_i)]_{ij}). \end{array} \tag{5.49}$$

και $0 < \epsilon_{\mathbf{U}_i} < 1$ μια προκαθορισμένη ανοχή, η τιμή της οποίας επηρεάζει το πλήθος των επαναλήψεων.

Τα υποπροβλήματα (5.42) ελαχιστοποιούνται επαναληπτικά μέχρις η καθολική συνθήκη σύγκλισης ικανοποιηθεί, η οποία στην ουσία ελέγχει την στασιμότητα της N-αδας των λύσεων $\mathbf{U}_i \forall i = 1, 2, \ldots, N$ όπου έχουν προκύψει. Η καθολική συνθήκη σύγκλισης ορίζεται ως εξής:

$$\|\nabla\varphi_1(\mathbf{U}_1{}^{(t)})\|_F + \ldots + \|\nabla\varphi_N(\mathbf{U}_N{}^{(t)})\|_F \leqslant \epsilon(\|\nabla\varphi_1(\mathbf{U}_1{}^{(1)})\|_F + \ldots + \|\nabla\varphi_N(\mathbf{U}_N{}^{(1)})\|_F) \tag{5.50}$$

Αλγόριθμος NTF Χρησιμοποιώντας την Ακολουθιακή κατά Συνιστώσα Μέθοδο

Σε αυτή την ενότητα παρουσιάζεται ένας αλγόριθμος παραγοντοποίησης μη αρνητικού τανυστή χρησιμοποιώντας την ακολουθιακή κατά συνιστώσα (sequential coordinate-wise) [44] μέθοδο επίλυσης NNLS προβλημάτων.

Ο αλγόριθμος PL-NTF βρίσκει λύση στο πρόβλημα της παραγοντοποίησης μη-αρνητικού τανυστή N-τάξης επιλύοντας ακολουθιακά τα υποπροβλήματα που ορίζονται στην (5.42) χρησιμοποιώντας την μέθοδο Projected Landweber. Την ίδια λογική ακολουθεί και ο αλγόριθμος CW-NTF. Αρχικά οι N πίνακες $\mathbf{U}^{(j)}$ αρχικοποιούνται με τυχαίες μη αρνητικές τιμές και τα υποπροβλήματα στην (5.42) επιλύονται επαναληπτικά μέχρις ότου ικανοποιηθεί η συνθήκη στασιμότητας στην (5.48) ή το πλήθος των επαναλήψεων υπερβεί ένα προκαθορισμένο αριθμό. Η διαφορά του CW-NTF από τον PL-NTF είναι η μέθοδος που ακολουθείται για την επίλυση των υποπροβλημάτων.

Οπώς έχει ήδη αναφερθεί κάθε υποπρόβλημα αποτελείται από ένα σύνολο προβλημάτων ελάχιστων μη αρνητικών τετραγώνων που ορίζεται στην (5.43). Κάθε τέτοιο πρόβλημα μπορεί να εκφραστεί ισοδύναμα ως τετραγωνικό πρόβλημα (Quadratic Problem - QP) ως εξής:

$$\min_{u_i^j \geqslant 0} \psi(\mathbf{u}_i^j), \ s.\ t.\ \mathbf{u}_i^j \geqslant 0. \tag{5.51}$$

όπου

$$\psi(\mathbf{u}_i^j) = \frac{1}{2}\mathbf{u}_i^j \mathbf{H} \mathbf{u}_i^{jT} + \mathbf{f}_i^{jT} \mathbf{u}_i^{jT}. \tag{5.52}$$

όπου $\mathbf{H} = \mathbf{Z}^T\mathbf{Z}$ και $\mathbf{f}_i^j = -\mathbf{Z}^T\mathbf{d}_i^{jT}$, και ο πίνακας $\mathbf{Z}$ ορίζεται όπως στην 5.40. Η ακολουθιακή κατά συνιστώσα μέθοδος ανανεώνει τα στοιχεία του $\mathbf{u}_i^j$ ένα ένα επαναληπτικά. Για λόγους απλοποίησης των συμβολισμών θέτουμε $\mathbf{u} = \mathbf{u}_i^{jT}$ και $\mathbf{d} = \mathbf{d}_i^{jT}$ έστω u_p η p-οστή συνιστώσα

του ανύσματος $\mathbf{u} = [u_1, u_2, \ldots, u_{I_i}]^T \in \mathbb{R}_+^{I_i}$, $\mathcal{I} = \{1, 2, \ldots, I_i\}$ και $\mathcal{I}_p = \mathcal{I} \setminus \{p\}$ Η συνάρτηση (5.52) μπορεί να εκφραστεί ως:

$$\psi(\mathbf{u}) = \frac{1}{2}\mathbf{u}^T\mathbf{H}\mathbf{u} + \mathbf{f}^T\mathbf{u}. \tag{5.53}$$

Η εξίσωση 5.53 έχει την κλασσική μορφή τετραγωνικής συνάρτησης και μπορεί να ξαναγραφεί ως:

$$\begin{aligned}
\psi(\mathbf{u}) &= \frac{1}{2}\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{I}} u_i u_j [\mathbf{H}]_{ij} + \sum_{i\in\mathcal{I}} u_i f_i \\
&= \frac{1}{2}u_p^2[\mathbf{H}]_{pp} + x_p f_p + u_p \sum_{i\in\mathcal{I}_p} u_i [\mathbf{H}]_{ik} + \sum_{i\in\mathcal{I}_p} u_i f_i \\
&+ \frac{1}{2}\sum_{i\in\mathcal{I}_p}\sum_{j\in\mathcal{I}_p} u_i u_j [\mathbf{H}]_{ij} \\
&= \frac{1}{2}u_p^2\alpha + u_p\beta + \gamma.
\end{aligned} \tag{5.54}$$

όπου

$$\alpha = [\mathbf{H}]_{pp} \tag{5.55}$$

$$\beta = f_p + \frac{1}{2}\sum_{i\in\mathcal{I}_p} u_i [\mathbf{H}]_{ip} = [\mathbf{H}\mathbf{u} + \mathbf{f}]_p - [\mathbf{H}]_{pp} u_p \tag{5.56}$$

$$\gamma = \sum_{i\in\mathcal{I}_p} u_i f_i + \frac{1}{2}\sum_{i\in\mathcal{I}_p}\sum_{j\in\mathcal{I}_p} u_i u_j [\mathbf{H}]_{ij}. \tag{5.57}$$

Το πρόβλημα της ελαχιστοποίησης της $\psi(\mathbf{u})$ ως προς την μεταβλητή u_p έχει αναλυτική λύση η οποία και είναι:

$$\begin{aligned}
u_p^* &= \min\ \psi(\mathbf{x}) \\
&= \min\ \frac{1}{2}u_p^2\alpha + u_p\beta + \gamma \\
&= \max(0, -\frac{\beta}{\gamma}) \\
&= \max(0, u_p - \frac{[\mathbf{H}\mathbf{u} + \mathbf{f}]_p}{[\mathbf{H}]_{pp}}).
\end{aligned} \tag{5.58}$$

Ο αλγόριθμος παραγοντοποίησης μη-αρνητικών τανυστών που χρησιμοποιέι την ακολουθιακή κατά συνιστώσα μέθοδο, ενημερώνει μόνο μια μεταβλητή σε κάθε επαναληπτικό βήμα. Ακολουθώντας την παραπάνω διαδικασία επαναληπτικά ενημερώνονται τα στοιχεία των N πινακών $\mathbf{U}^{(j)}$ και η διαδικασία σταματά όταν η N-αδα των λύσεων $\mathbf{U}_i, i = 1, 2, \ldots, N$ αποτελέι στασιμό σημείο της (5.38) ή είναι επαρκώς κοντά σε αυτό. Μπορεί και εδώ να χρησιμοποιηθεί η καθολική συνθήκη ελέγχου στασιμότητας που ορίζεται από την (5.50). Ο Franc στο [44] εγγυάται μαθηματικά την σύγκλιση του αλγορίθμου σε στάσιμο σημείο.

Κεφάλαιο 6

Πειραματική Αξιολόγηση

Στο παρόν κεφάλαιο περιγράφεται η πειραματική διαδικασία που εφαρμόστηκε στη διατριβή με στόχο την αξιολόγηση της προτεινόμενης βίο-εμπνευσμένης πολυγραμμικής προσέγγισης στο πρόβλημα της αναγνώρισης μουσικού είδους. Στην ενότητα 6.1 παρουσιάζεται η διαδικασία κατασκευής των τανυστών δεδομένων από τις αναπαραστάσεις φλοιού. Στην ενότητα 6.2 περιγράφεται η διαδικασία εξαγωγής χαρακτηριστικών διανυσμάτων από τις αναπαραστάσεις φλοιού με την χρήση πολυγραμμικών τεχνικών ανάλυσης υποχώρων. Τα αποτελέσματα των πειραμάτων και τα ποσοστά ορθής ταξινόμησης μουσικού έιδους παρουσιάζονται στην ενότητα 6.3. Τέλος, στην ενότητα 6.4 αναφέρονται τα τελικά συμπεράσματα καθώς και πιθανές μελλοντικές ερευνητικές κατευθύνσεις.

6.1 Δημιουργία Τανυστών Δεδομένων

Για την αξιολόγηση των αναπαραστάσεων φλοιού ως φέροντα πληροφορίας για το μουσικό είδος εκτελούνται πειράματα σε δύο σύνολα δεδομένων, στο σύνολο δεδομένων GTZAN, που δημιουργήθηκε από τον Τζανετάκη [126] και στο σύνολο δεδομένων ISMIR2004GENRE το οποίο χρησιμοποιήθηκε για την αξιολόγηση αλγορίθμων αναγνώρισης μουσικού είδους στα πλαίσια του συνεδρίου ISMIR 2004.

Όπως έχει ήδη αναφερθεί στην ενότητα 4.2.1, η βάση GTZAN περιέχει 1000 αρχεία, καλύπτοντας 10 μουσικά είδη: Classical, Country, Disco, HipHop, Jazz, Rock, Blues, Reggae, Pop, και Metal. Σε κάθε κλάση ανήκουν 100 αρχεία. Κάθε αρχείο έχει κατάληξη .au και έχει διάρκεια περίπου 30 sec. Τα αρχεία είναι μονοφωνικά, με συχνότητα δειγματοληψίας στα 22.050 Hz, με 16 bits ανά δείγμα. Για την δημιουργία του αναπαραστάσεων φλοιού από τα αρχεία του συνόλου δεδομένων, τα αρχεία υπόκεινται προ-επεξεργασία. Αρχικά κάθε αρχείο υποβάλλεται σε υποδειγματολειψία στα 8 kHz, εν συνεχεία το ψηφιακό σήμα κανονικοποιείται αφαιρώντας

την μέση τιμή του και διαιρώντας με την τυπική του απόκλιση.

Το σύνολο δεδομένων ISMIR2004GENRE αποτελείται από 1458 πλήρη αρχεία μουσικής, κατανεμημένα ανομοιογενώς σε 6 μουσικά είδη classical, electronic, jazzblues, metalpunk, rockpop, world. Κάθε αρχείο έχει κατάληξη .mp3 και είναι στερεοφωνικό, με συχνότητα δειγματοληψίας στα 44.1 kHz, με 16 bits ανά δείγμα. Για την δημιουργία του αναπαραστάσεων φλοιού από τα αρχεία του συνόλου δεδομένων, τα αρχεία μετατρέπονται σε μονοφωνικά και σε format .wav. Από κάθε αρχείο επιλέγονται τα 30 sec. που ακολουθούν τα πρώτα 30 sec. του μουσικού κομματιού. Αυτό γίνεται διότι στα πρώτα 30 sec. του κομματιού είναι πιθανόν να υπάρχουν τμήματα παύσεων, εισαγωγικά μουσικά μοτίβα κ.α. που δεν αποτελούν χαρακτηριστικά του μουσικού είδους. Εν συνεχεία, το επιλεγμένο σήμα διάρκειας 30 sec. υποβάλλεται σε υποδειγματολειψία στα 8 kHz, και κανονικοπιείται αφαιρώντας την μέση τιμή του και διαιρώντας με την τυπική απόκλιση του. Η κανονικοποίηση του σήματος είναι μείζονος σημασίας, διότι έτσι μετριάζουμε το *φαινόμενο του παραγωγού* (producer effect) [97], εφόσον αυτό υπάρχει στα δεδομένα.

Για την αξιολόγηση του προτεινόμενου πλαισίου αναγνώρισης μουσικού είδους χρησιμοποιούμε διαστρωματωμένη διασταυρωμένη επικύρωση 10 - βημάτων (10-fold cross validation) κατά την πειραματική διαδικασία. Σε κάθε βήμα της διασταυρωμένης επικύρωσης, καθένα από τα σύνολα δεδομένων που χρησιμοποιούνται χωρίζεται σε δύο υποσύνολα, το ένα χρησιμοποιείται για εκπαίδευση (training) ενώ το άλλο για έλεγχο (testing). Όπως αναφέρθηκε στο κεφάλαιο 2 κάθε ηχογράφηση αναπαρίσταται είτε από έναν τανυστή τρίτης τάξης $\mathcal{D} \in \mathbb{R}_{+}^{I_{scales} \times I_{rates} \times I_{frequencies}}$ εάν πρόκειται για την από κοινού χρονοφασματική αναπαράσταση φλοιού διαμορφώσεων του σήματος είτε από ένα τανυστή δεύτερης τάξης $\mathcal{D} \in \mathbb{R}_{+}^{I_{frequencies} \times \times I_{modulationfrequencies}}$ εάν πρόκειται για την αναπαράσταση φλοιού χρονικών διαμορφώσεων. Συνεπώς, στοιβάζοντας τους τανυστές φλοιωδών αναπαραστάσεων κάθε ηχογράφησης δημιουργείται ένας τανυστής δεδομένων εκπαίδευσης τέταρτης τάξης $\mathcal{T}_{stm} \in \mathbb{R}_{+}^{I_{samples} \times I_{scales} \times I_{frequencies}}$ που περιλαμβάνει τους τανυστές χρονοφασματικών διαμορφώσεων και ένας τανυστής δεδομένων εκπαίδευσης τρίτης τάξης που περιλαμβάνει τους τανυστές χρονικών διαμορφώσεων $\mathcal{T}_{tm} \in \mathbb{R}_{+}^{I_{samples} \times I_{frequencies} \times I_{modulationfrequencies}}$ όπου *samples* είναι το πλήθος των ηχογραφήσεων σε κάθε σύνολο εκπαίδευσης.

6.2 Εξαγωγή Χαρακτηριστικών από τις Αναπαραστάσεις Φλοιού

Για την εξαγωγή χαρακτηριστικών διανυσμάτων από τις αναπαραστάσεις φλοιού που είναι κατάλληλα για ταξινόμηση χρησιμοποιούνται οι πολυγραμμικές τεχνικές MPCA, HOSVD, και NTF. Στις παραγράφους που ακολουθούν περιγράφεται η διαδικασία εξαγωγής χαρακτηριστικών διανυσμάτων από τις αναπαραστάσεις φλοιού που περιγράφουν τις από κοινού χρονοφασματικες διαμορφώσεις του ακουστικού σήματος. Η εξαγωγή χαρακτηριστικών από τις αναπαραστάσεις φλοιού που περιγράφουν τις χρονικές διαμορφώσεις πραγματοποιείται με όμοιο τρόπο.

Εξαγωγή Χαρακτηριστικών με την HOSVD

Ο τανυστής εκπαίδευσης $\mathcal{T}_{stm}$ αποσυντίθεται στα mode-n ιδιάζοντα διανύσματα χρησιμοποιώντας τον αλγόριθμο που παρουσιάστηκε στην ενότητα 5.3.3. Από τα ιδιάζοντα διανύσματα κατασκευάζονται οι πίνακες $\mathbf{U}^{(scales)}$, $\mathbf{U}^{(rates)}$, $\mathbf{U}^{(frequencies)}$. Οι πίνακες αυτοί είναι ορθοκανονικοί και διατεταγμένοι κατ' αναλογία με την ανάλυση ιδιαζουσών τιμών. Για να παράγουμε έναν υποχώρο που προσεγγίζει τον αρχικό χώρο δεδομένων από κάθε πίνακα ιδιαζουσών διανυσμάτων κρατάμε τα πρώτα p ιδιάζοντα διανύσματα έτσι ώστε να διατηρηθούν μόνο οι επιθυμητοί κύριοι άξονας σε κάθε mode.

Κάθε τανυστής χαρακτηριστικών $\mathcal{D}_i$ που αντιστοιχεί στην i-οστη ηχογράφηση του συνόλου δεδομένων προβάλλεται επάνω σε αυτούς τους αποκομμένους ορθοκανονικούς άξονες $\tilde{\mathbf{U}}^{(scales)}$, $\tilde{\mathbf{U}}^{(rates)}$, $\tilde{\mathbf{U}}^{(frequencies)}$ και έτσι προκύπτει ένας τανυστής χαρακτηριστικών μικρότερων διαστάσεων: $\tilde{\mathcal{X}}_i$.

$$\tilde{\mathcal{X}}_i = \mathcal{D}_i \times_1 \tilde{\mathbf{U}}^{(scales)} \times_2 \tilde{\mathbf{U}}^{(rates)} \times_3 \tilde{\mathbf{U}}^{(frequencies)}. \tag{6.1}$$

Τα χαρακτηριστικά διανύσματα που είναι κατάλληλα για ταξινόμηση $\tilde{\mathbf{x}}_i$ προκύπτουν από την μετατροπή του τανυστή $\tilde{\mathcal{X}}_i$ σε διάνυσμα.

Εξαγωγή Χαρακτηριστικών με την MPCA

Παρομοίως με την HOSVD, ένας πολυγραμμικός μετασχηματισμός τέτοιος ώστε να απεικονίζει τον αρχικό τανυστικό χώρο $\mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ σε έναν τανυστικό υποχώρο $\mathbb{R}^{P_1 \times P_2 \times \ldots \times P_N}$ με $P_n < I_n$, $\forall n$ ο οποίος υποχώρος διατηρεί την μεταβλητότητα του αρχικού τανυστικού χώρου στον οποίο ορίζεται ο $\mathcal{T}_{stm}$, προκύπτει εφαρμόζοντας τον αλγόριθμο που περιγράφεται στην ενότητα 5.3.2. Κάθε τανυστής χαρακτηριστικών $\mathcal{D}_i$ που αντιστοιχεί στην i-οστη ηχογράφηση του συνόλου δεδομένων προβάλλεται στον μετασχηματισμένο τανυστικό χώρο που προέκυψε από

την MPCA.Τα χαρακτηριστικά διανύσματα που είναι κατάλληλα για ταξινόμηση $\tilde{\mathbf{x}}_i$ προκύπτουν από την μετατροπή του τανυστή $\tilde{\mathcal{X}}_i$ σε διάνυσμα.

Εξαγωγή Χαρακτηριστικών με την NTF

Τρεις αλγόριθμοι παραγοντοποιήσης μη αρνητικών τανυστών χρησιμοποιούνται για την εξαγωγή χαρακτηριστικών από τις αναπαραστάσεις φλοιού του ήχου. Η διαδικασία εξαγωγής χαρακτηριστικών είναι κοινή κα για τους τρεις αλγορίθμους και ακολουθεί.

Ο τανυστής δεδομένων $\mathcal{T}_{stm}$ προσεγγίζεται ως άθροισμα k rank-1 τανυστών οι οποίοι προκύπτουν από κάποιον αλγόριθμο NTF. Χωρίς βλάβη της γενικότητας, αυτό εκφράζεται σε μορφή πινάκων ως εξής:

$$\begin{aligned} \mathbf{T}_{stm(1)} &= \mathbf{U}^{(1)}(\mathbf{U}^{(4)} \odot \mathbf{U}^{(3)} \odot \mathbf{U}^{(2)})^T \Longleftrightarrow \\ \mathbf{T}^T_{stm(1)} &= (\mathbf{U}^{(4)} \odot \mathbf{U}^{(3)} \odot \mathbf{U}^{(2)})\mathbf{U}^{(1)T}, \end{aligned} \tag{6.2}$$

όπου $\mathbf{T}_{stm(1)} \in \mathbb{R}_+^{samples \times (scales \cdot rate \cdot frequencies)}$ είναι το mode-1 ανάπτυγμα του τανυστή εκπαίδευσης $\mathcal{T}_{stm}$, $\mathbf{U}^{(1)} \in \mathbb{R}_+^{samples \times k}$,$\mathbf{U}^{(2)} \in \mathbb{R}_+^{scales \times k}$, $\mathbf{U}^{(3)} \in \mathbb{R}_+^{rates \times k}$, και $\mathbf{U}_+^{(4)} \in \mathbb{R}^{frequencies \times k}$.

Από την Εξίσωση (6.2), είναι εμφανές ότι κάθε στήλη του $\mathbf{T}^T_{stm(1)}$, δηλαδή η ανηγμένη σε διάνυσμα αναπαράσταση φλοιού, αποτελεί το γραμμικό συνδυασμό συναρτήσεων βάσης από τον πίνακα $\mathbf{W} = \mathbf{U}^{(4)} \odot \mathbf{U}^{(3)} \odot \mathbf{U}^{(2)}$ με συντελεστές από τον πίνακα συντελεστων $\mathbf{U}^{(1)T}$. Εφαρμόζοντας την διαδικασία Gram-Schmidt στον πίνακα βάσεων $\mathbf{W}$, προκύπτει ένας ορθοκανονικός πίνακας βάσεων $\mathbf{Q}$. Ο πίνακας $\mathbf{Q}$ παράγει τον ίδιο υποχώρο με τον $\mathbf{W}$. Η διαδικασία της ορθκανονικοποιήσης υιοθετήθηκε διότι αποτελέσματα ερευνών [22, 40] ότι η ορθογωνιότητα αυξάνει την διακριτική ισχύ της προβολής. Συνεπώς τα χαρακτηριστικά διανύσματα που είναι κατάλληλα για ταξινόμηση προκύπτουν ως $\tilde{\mathbf{x}}_i = \mathbf{Q}^T\mathbf{d}_i$, όπου $\mathbf{d}_i$ είναι η ανηγμένη σε διάνυσμα αναπαράσταση φλοιού της i-οστής ηχογράφησης του συνόλου δεδομένων.

6.3 Πειραματικά Αποτελέσματα

Στην παρούσα μεταπτυχιακή διατριβή εξάγονται χαρακτηριστικά διανύσματα από τις δυο αναπαραστάσεις φλοιού που παρουσιάστηκαν στο κεφάλαιο 2, σύμφωνα με τη διαδικασία που περιγράφεται στην ενότητα 6.2. Για την αξιολόγηση των χαρακτηριστικών και των πολυγραμμικών μεθόδων ανάλυσης υποχώρων εκτελούνται εκτεταμένα πειράματα στα σύνολα δεδομένων GTZAN και ISMIR2004GENRE. Στα πειράματα που ακολουθούν, ως ταξινομητές χρησιμοποιούνται η *Μηχανή Εδραίων Διανυσμάτων* (Support Vector Machine - SVM)[128] και ο ταξινομητής *Πλησιέστερου Γείτονα* (Nearest Neighbour - NN). Για την Μηχανή Εδραίων Διανυσμάτων χρησιμοποιήθηκε πυρήνας ακτινικής βάσης (RBF) καθώς και γραμμικός πυρήνας

Απόσταση	L_1	L_2	CSM
$d(\mathbf{a},\mathbf{b})$	$\sum_{h=1}^{H} \lvert\mathbf{a}(h)-\mathbf{b}(h)\rvert$	$\sqrt{\sum_{h=1}^{H}[\mathbf{a}(h)-\mathbf{b}(h)]^2}$	$-\frac{\sum_{h=1}^{H}\mathbf{a}(h)\mathbf{b}(h)}{\sqrt{\sum_{h=1}^{H}\mathbf{a}(h)^2\sum_{h=1}^{H}\mathbf{b}(h)^2}}$

Πίνακας 6.1: Τρία μέτρα απόστασης τα οποία δοκιμάστηκαν για την αναγνώριση μουσικού είδους.

(Linear). Οι παράμετροι του RBF πυρήνα προσδιορίζονται από έναν εξαντλητικό αλγόριθμο αναζήτησης, παρόμοιο με αυτόν που περιγράφεται στο [60]. Για την ταξινόμηση μουσικού είδους με τον ταξινομητή πλησιέστερου γείτονα χρησιμοποιούμε τρία διαφορετικά μέτρα απόστασης που συνοψίζονται στον πίνακα 6.1. Για κάθε διάνυσμα χαρακτηριστικών ελέγχου $\mathbf{a} = \tilde{\mathbf{x}}_i$, υπολογίζεται η απόσταση του με κάθε διάνυσμα εκπαίδευσης $\mathbf{b} = \tilde{\mathbf{x}}_j$, έστω η απόσταση $d(\mathbf{a},\mathbf{b})$, το διάνυσμα ταξινομείται στην κλάση για την οποία το μέτρο της απόστασης έχει την μεγαλύτερη τιμή. Στις παραγράφους που ακολουθούν παρουσιάζονται τα αποτελέσματα των πειραμάτων χρησιμοποιώντας χαρακτηριστικά που εξάγονται από τις αναπαραστάσεις φλοιού του ήχου και για τα δύο σύνολα δεδομένων.

6.3.1 Πειραματικά Αποτελέσματα Χαρακτηριστικών Χρονικών και Συχνοτικών Διαμορφώσεων

Στους πίνακες 6.2 και 6.3 παρουσιάζεται τα η ακρίβεια ορθής ταξινόμησης μουσικού είδους για το σύνολο δεδομένων GTZAN και ISMIR2004GENRE αντίστοιχα, με χαρακτηριστικά που εξήχθησαν από την από κοινού χρονοφασματική αναπαράσταση διαμορφώσεων χρησιμοποιώντας τους τρεις αλγορίθμους NTF, την HOSVD και την MPCA. Κάθε γραμμή του εκάστοτε πίνακα προσδιορίζει την πολυγραμμική μέθοδο ανάλυσης υποχώρων που χρησιμοποιήθηκε για την εξαγωγή των χαρακτηριστικών διανυσμάτων. Με PL-NTF υποδηλώνεται η προτεινομένη μέθοδος παραγοντοποιήσης μη αρνητικών τανυστών με την χρήση του αλγορίθμου Landweber. Με CW-NTF υποδηλώνεται ο προτεινόμενος Coordinate Wise NTF αλγόριθμος, ενώ με NTF υποδηλώνεται αλγόριθμος που προτάθηκε από τον Μπενέτο στο [8, 9], και χρησιμοποιεί την νόρμα Frobenius. Οι στήλες του εκάστοτε πίνακα προσδιορίζουν τον ταξινομητή που χρησιμοποιήθηκε για την αναγνώριση του μουσικού είδους. Για την μηχανή εδραίων διανυσμάτων (SVM) οι ετικέτες RBF και Linear προσδιορίζουν τον τύπο του πυρήνα που χρησιμοποιήθηκε, ακτινικής βάσης και γραμμικό αντίστοιχα. Όσον αφορά τις ετικέτες L1, L2 και CSM υποδηλώνουν το μέτρο απόστασης που χρησιμοποιήθηκε από τον ταξινομητή πλησιέστερου γείτονα (NN) και αντιστοιχούν στις νόρμες L_1, L_2 και στο μέτρο ομοιότητας συνημίτονου. Τα αποτελέσματα που παρουσιάζονται αποτελούν την μέση ορθή ταξινόμηση που προκύπτει από τα αποτελέσματα της διασταυρωμένης επικύρωσης. Οι τιμές που βρίσκονται σε παρενθέσεις δηλώνουν την τυπική

Σύνολο Δεδομένων:GTZAN					
	SVM		**NN**		
	RBF	Linear	L_1	L_2	CSM
PL-NTF	79.70%(2.05)	**75.80%**(3.64)	63.50%(3.50)	63.30%(2.71)	64.20%(2.78)
CW-NTF	**80.90%**(1.96)	71.50%(4.08)	61.50%(5.19)	**63.77%**(5.42)	**64.90%**(3.54)
NTF	79.50%(2.41)	73.30%(2.94)	64.30%(2.94)	61.50%(3.30)	64.40%(3.56)
HOSVD	77.50%(4.30)	71.03%(8.54)	**64.80%**(3.73)	61.70%(3.83)	61.50%(4.55)
MPCA	75.02%(4.49)	71.50%(4.37)	63.00%(4.02)	61.20%(4.21)	61.50%(4.14)

Πίνακας 6.2: Αποτελέσματα ορθής ταξινόμησης με χρήση χαρακτηριστικών από κοινού συχνοτικής και χρονικής διαμορφώσης στο σύνολο δεδομένων GTZAN.

Σύνολο Δεδομένων:ISMIR2004GENRE					
	SVM		**NN**		
	RBF	Linear	L_1	L_2	CSM
PL-NTF	**81.48%**(2.41)	73.84%(3.13)	**73.37%**(1.85)	73.17%(2.27)	**72.96%**(2.66)
CW-NTF	80.34%(2.45)	**73.94%**(2.50)	72.76%(2.20)	72.28%(2.49)	72.29%(2.34)
NTF	80.47%(2.26)	72.54%(2.49)	71.29%(3.37)	**73.37%**(3.04)	71.26%(3.51)
HOSVD	77.67(3.19)%	72.76%(3.37)	71.54%(2.53)	70.26%(3.16)	70.60%(3.37)
MPCA	78.23(2.52)%	70.66%(3.91)	72.15%(2.75)	71.34%(3.31)	73.33%(3.75)

Πίνακας 6.3: Αποτελέσματα ορθής ταξινόμησης με χρήση χαρακτηριστικών από κοινού συχνοτικής και χρονικής διαμορφώσης στο σύνολο δεδομένων ISMIR2004GENRE.

απόκλιση.

Οι τιμές των παραμέτρων που χρησιμοποιήθηκαν στα πειράματα έχουν ως εξής: για τους τρεις NTF η τιμή της τάξης επιλέχτηκε πειραματικά και είναι $k = 180$ για το σύνολο δεδομένων GTZAN και $k = 170$ για το σύνολο δεδομένων ISMIR2004GENRE. Ο αριθμός των κύριων αξόνων που προκύπτουν από την HOSVD είναι 5 από 6 για την διάσταση *rate*, 7 από 10 για την διάσταση *scale* και 12 από 128 για την διάσταση των συχνοτήτων και για τα δύο σύνολα δεδομένων. Τέλος τα χαρακτηριστικά που εξάγονται με τη χρήση της MPCA περιλαμβάνουν το 98% της συνολικής μεταβλητότητας σε κάθε διάσταση του τανυστή δεδομένων, επίσης και για τα δύο σύνολα δεδομένων.

Σύνολο Δεδομένων: GTZAN					
	SVM		**NN**		
	RBF	Linear	L_1	L_2	CSM
PL-NTF	81.4%(3.16)	75.1%(3.57)	72.3%(4.29)	73.7%(3.77)	**72.6%**(3.83)
CW-NTF	**82.6%**(3.97)	**76.9%**(3.69)	72.1%(4.62)	**73.8%**(4.13)	71.4%(3.68)
NTF	80.1%(2.92)	73.6%(2.83)	**72.6%**(3.06)	73.3%(3.88)	**72.6%**(3.83)
HOSVD	71.40%(2.17)	66.1%(2.88)	66.2%(3.91)	69.1%(3.54)	68.2%(4.61)
MPCA	73.70%(3.59)	69.5%(3.06)	67.40%(3.59)	67.9%(2.84)	66.6%(2.95)

Πίνακας 6.4: Αποτελέσματα ορθής ταξινόμησης με χρήση χαρακτηρηστικών χρονικών διαμορφώσεων στο σύνολο δεδομένων GTZAN.

6.3.2 Πειραματικά Αποτελέσματα Χαρακτηριστικών Χρονικών Διαμορφώσεων

Κατά αναλογία με την ενότητα 6.3.1, στους πίνακες 6.4 και 6.5 παρουσιάζεται τα η ακρίβεια ορθής ταξινόμησης μουσικού είδους για το σύνολο δεδομένων GTZAN και ISMIR2004GENRE αντίστοιχα, με χαρακτηριστικά που εξήχθησαν από την αναπαράσταση χρονικών διαμορφώσεων χρησιμοποιώντας τους τρεις αλγορίθμους NTF, την HOSVD και την MPCA.

Οι τιμές των παραμέτρων που χρησιμοποιήθηκαν στα πειράματα έχουν ως εξής: για τους τρεις NTF η τιμή της τάξης επιλέχτηκε πειραματικά και είναι $k = 200$ για το σύνολο δεδομένων GTZAN και $k = 180$ για το σύνολο δεδομένων ISMIR2004GENRE. Ο αριθμός των κύριων αξόνων που προκύπτουν από την HOSVD είναι 5 από 8 για την διάσταση των συχνότητων χρονικής διαμόρφωσης, 61 από 96 για την διάσταση των συχνοτήτων και για τα δύο σύνολα δεδομένων. Τέλος τα χαρακτηριστικά που εξάγονται με τη χρήση της MPCA περιλαμβάνουν το 98% της συνολικής μεταβλητότητας σε κάθε διάσταση του τανυστή δεδομένων, και για τα δύο σύνολα δεδομένων.

6.3.3 Σχολιασμός Πειραματικών Αποτελεσμάτων

Με βάση τα αποτελέσματα που παρουσιάζονται στους Πίνακες 6.2 - 6.5 συμπεραίνουμε ότι οι αλγόριθμοι παραγοντοποιήσης μη αρνητικών τανυστών εξάγουν εν γένει χαρακτηριστικά διανύσματα με μεγαλύτερη διακριτική ισχύ σε σύγκριση με εκείνα που εξάγονται από την HOSVD και από την MPCA και από τις δυο αναπαραστάσεις φλοιού. Η μηχανή εδραίων διανυσμάτων με RBF πυρήνα κρίνεται ως ο πιο αποδοτικός αλγόριθμος ταξινόμησης για τα μη αρνητικά χαρακτηριστικά και για τα δύο σύνολα δεδομένων, ξεπερνώντας την ακρίβεια τόσο της μηχανής

Σύνολο Δεδομένων:ISMIR2004GENRE					
	SVM		**NN**		
	RBF	Linear	L_1	L_2	CSM
PL-NTF	83.16%(2.16)	73.83%(2.79)	75.73%(2.53)	**76.88%**(3.30)	**78.10%**(2.43)
CW-NTF	**83.57%**(3.57)	**75.19%**(3.21)	**76.95%**(2.96)	76.25%(3.79)	77.53%(3.72)
NTF	83.03%(3.33)	72.36%(2.99)	76.34%(3.95)	76.34%(3.51)	76.95%(3.02)
HOSVD	77.90%(2.62)	71.50%(2.99)	72.90%(2.48)	73.30%(3.23)	73.30%(2.68)
MPCA	78.68%(1.87)	72.22%(3.55)	72.67%(2.21)	73.90%(3.45)	77.83%(2.52)

Πίνακας 6.5: Αποτελέσματα ορθής ταξινόμησης με χρήση χαρακτηρηστικών χρονικών διαμορφώσεων στο σύνολο δεδομένων ISMIR2004GENRE.

εδραίων διανυσμάτων με γραμμικό πυρήνα όσο και του ταξινομητή πλησιέστερου γείτονα.

Επιπλέον, παρατηρούμε ότι τα χαρακτηριστικά διανύσματα που προκύπτουν από τους δύο προτεινόμενους αλγόριθμους LP-NTF και CW-NTF υπερτερούν οριακά αυτών που προκύπτουν από τον πιο κλασικό αλγόριθμο NTF. Το γεγονός αυτό εικάζεται ότι οφείλεται στις ιδιότητες σύγκλισης των προτεινόμενων αλγορίθμων.

Είναι φανερό ότι ο ταξινομητής πλησιέστερου γείτονα συμπεριφέρεται πιο αποδοτικά στο σύνολο δεδομένων ISMIR2004GENRE από ότι στο σύνολο δεδομένων GTZAN, πράγμα που πιθανόν να οφείλεται στο φαινόμενο του παραγωγού [97], μια και στα πειράματα δεν έχει ληφθεί μέριμνα έτσι ώστε μουσικά κομμάτια από τον ίδιο δίσκο-καλλιτέχνη να ανήκουν αποκλειστικά και μόνο στο σύνολο εκπαίδευσης ή στο σύνολο ελέγχου, όπως προτείνει ο Pampalk. Μία άλλη εξήγηση που μπορεί να δοθεί είναι ότι το σύνολο δεδομένων ISMIR2004GENRE περιέχει ένα σημαντικό αριθμό κομματιών κλασσικής μουσικής που γενικά ως μουσικό είδος διαφέρει αρκετά από τα υπόλοιπα του συνόλου, οπότε και είναι ευκολότερο να αναγνωρισθεί.

Συγκρίνοντας τις αναπαραστάσεις φλοιού ως πηγές χαρακτηριστικών διαπιστώνουμε ότι η πληροφορία μόνο των χρονικών διαμορφώσεων έχει μεγαλύτερη διακριτική ισχύ από την πληροφορία που προκύπτει από τις από κοινού χρονοφασματικές διαμορφώσεις.

Η υψηλότερη ακρίβεια ορθής ταξινόμησης για το σύνολο δεδομένων GTZAN προκύπτει από χαρακτηριστικά που εξάγονται με τον CW-NTF και ταξινομούνται με τη χρήση SVM με RBF πυρήνα. Το ποσοστό ορθής ταξινόμησης είναι ίσο με 82.6% και ξεπερνά τα ποσοστό ορθής ταξινόμησης που αναφέρεται από τον Τζανετάκη στο [126] και είναι ίσο με 61.0%, τον Lidy στο [79] και ισούται με 76.8%, του Holzapfel στο [56] που ισούται με 74%, του Li στο [76] που ισούται με 78.5% και του Μπενέτου [8, 9] που έιναι ίση με 75%. Η επίδοση συγκρίνεται με αυτή που αναφέρεται από τον Bergstra στο [12] και ισούται με 82.5%. Να σημειωθεί ότι στο [12] δεν αναφέρεται αν το συγκεκριμένο ποσοστό έχει προκύψει από διασταυρωμένη επικύρωση.

Η υψηλότερη ακρίβεια ορθής ταξινόμησης για το σύνολο δεδομένων ISMIR2004GENRE προκύπτει από χαρακτηριστικά που εξάγονται με τον CW-NTF και ταξινομούνται με τη χρήση SVM με RBF πυρήνα και ισούται με 83.57%. Να σημειωθεί ότι και τα χαρακτηριστικά που εξάγονται χρησιμοποιώντας τους υπολοίπους NTF αλγόριθμους οδηγούν σε ακρίβεια ορθής ταξινόμησης μεγαλύτερης του 83%. Δεν είναι δυνατό να συγκρίνουμε τα προαναφερθέντα ποσοστά με ποσοστά που έχουν επιτευχθεί από άλλους αλγορίθμους αναγνώρισης μουσικού είδους στο συγκεκριμένο σύνολο δεδομένων λόγω των διαφορετικών πειραματικών διαδικασιών που ακολουθήθηκαν.

Από την παραπάνω ανάλυση των αποτελεσμάτων προκύπτει ότι είναι δυνατόν να εξαχθούν χαρακτηριστικά από τις αναπαραστάσεις φλοιού με σημαντική διακριτική ισχύ που έστω και οριακά ξεπερνά αυτή των κλασικών χαρακτηριστικών που έχουν προταθεί στην βιβλιογραφία και έχουν χρησιμοποιηθεί στο πρόβλημα της αναγνώρισης μουσικού είδους. Ικανοποιητική κρίνεται και η επίδοση που προκύπτει και από τα χαρακτηριστικά που προκύπτουν από την HOSVD και από την MPCA.

Παρόλα αυτά, στα πλαίσια της παρούσης μεταπτυχιακής, δεν κατέστη δυνατό να ξεπεραστεί ουσιαστικά το ποσοστό του 83% που αποτελεί ένα άνω όριο επίδοσης για όλα τα πρόσφατα συστήματα αυτόματης αναγνώρισης μουσικού είδους. Το γεγονός αυτό μπορεί να οφείλεται αφενός στην ποιότητα των δεδομένων που χρησιμοποιούνται για την αξιολόγηση των αλγορίθμων και αφετέρου ίσως το ποσοστό αυτό να αποτελεί το άνω όριο διακριτοποίησης των μουσικών ειδών χρησιμοποιώντας πληροφορίες που εξάγονται μόνο από το μουσικό σήμα.

6.4 Μελλοντικές Κατευθύνσεις

Σ' αυτή την μεταπτυχιακή διατριβή εξετάστηκε το πρόβλημα της αυτόματης αναγνώρισης μουσικού είδους υπό μια βιο-εμπνευσμένη πολυγραμμική οπτική. Χρησιμοποιήθηκαν υπολογιστικά μοντέλα του συστήματος ακοής ώστε να εξαχθούν βίο-εμπνευσμένες αναπαραστάσεις των μουσικών σημάτων, τις αποκαλούμενες φλοιώδεις αναπαραστάσεις. Από τις φλοιώδεις αναπαραστάσεις εξήχθησαν χαρακτηριστικά διανύσματα χρησιμοποιώντας τρεις κατηγορίες πολυγραμμικών τεχνικών ανάλυσης υποχώρων, την HOSVD, την MPCA και την NTF. Στα πλαίσια της διατριβής προτάθηκαν και δύο νέοι αλγόριθμοι παραγοντοποιήσης μη αρνητικών τανυστών N-οστής τάξης με ισχυρά θεμελιωμένες ιδιότητες σύγκλισης. Εκτεταμένα πειράματα σε δύο διαδεδομένα σύνολα δεδομένων, που έχουν χρησιμοποιηθεί στην βιβλιογραφία για την αξιολόγηση αλγορίθμων αυτόματης αναγνώρισης μουσικού είδους, κατέδειξαν την υπεροχή των χαρακτηριστικών χρονικών διαμορφώσεων έναντι αυτών των από κοινού χρονοφασματικών διαμορφώσεων. Επίσης τα αποτελέσματα των πειραμάτων δείχνουν την υπεροχή των NTF μεθόδων εν συγκρίσει

με τις άλλες πολυγραμμικές τεχνικές εξαγωγής χαρακτηριστικών. Η καλύτερη επίδοση που σημειώνεται στο σύνολο δεδομένων GTZAN είναι 82.6% και υπερτερεί των επιδόσεων όλων των αλγορίθμων που έχουν προταθεί στην βιβλιογραφία και έχουν ελεγχθεί στο συγκεκριμένο σύνολο δεδομένων. Για το σύνολο δεδομένων ISMIR2004GENRE η καλύτερη επίδοση που σημειώνεται είναι 83.57%, αλλά δεν μπορεί να συγκριθεί άμεσα με τις επιδόσεις άλλων αλγορίθμων αναγνώρισης μουσικού είδους στο ίδιο σύνολο δεδομένων λόγω των διαφορετικών πειραματικών διαδικασιών που ακολουθήθηκαν.

Οι πολυγραμμικές τεχνικές που εφαρμόστηκαν για την εξαγωγή χαρακτηριστικών είναι μη επιτηρούμενες, δηλαδή δεν ενσωματώνουν πληροφορία κλάσης. Μελλοντικά στόχος είναι να αναπτυχθούν επιτηρούμενες πολυγραμμικές τεχνικές ανάλυσης υποχώρων που βασίζονται στη παραγοντοποιήση μη αρνητικών τανυστών. Επιπλέον στόχος είναι να μελετηθούν σε βάθος οι παράμετροι των μοντέλων εξαγωγής φλοιωδών αναπαραστάσεων και πως αυτές επηρεάζουν την ποιότητα των αναπαραστάσεων. Μια ακόμη μελλοντική ερευνητική κατεύθυνση θα μπορούσε να περιλαμβάνει την αξιολόγηση της προτεινόμενης βίο-εμπνευσμένης πολυγραμμικής προσέγγισης σε άλλα συγγενή προβλήματα με αυτό της αναγνώρισης μουσικού είδους.

Τέλος, στη παρούσα εργασία, έχουμε υποθέσει ότι κάθε μουσικό κομμάτι ανήκει σε μια και μοναδική κλάση μουσικού είδους. Προφανώς είναι ρεαλιστικό να θεωρήσουμε ότι κάθε μουσικό κομμάτι ανήκει, πολλές φορές, σε περισσότερες από μια κατηγορίες μουσικού είδους, χαρακτηρίζοντας έτσι και το μουσικό στυλ του κομματιού [12]. Οι τανυστές υψηλής τάξης είναι εν γένει δομές που θα μπορούσαν να χρησιμοποιηθούν σε ένα τέτοιο multi-labelling πρόβλημα ταξινόμησης.

Βιβλιογραφία

[1] R. Albright, J. Cox, D. Duling, A. Langville, and C. D Meyer, "Algorithms, initializations, and convergence for the nonnegative matrix factorization", preprint, 2006.

[2] L. Atlas and S. Shamma, "Joint Acoustic and Modulation Frequency", *EURASIP Journal Applied Signal Processing*, Vol. 2003, No. 7, pp. 668-675, 2003.

[3] J. J. Aucouturier and F. Pachet, "Representing musical genre: a state of the art", *Journal New Music Research*, Vol. 32, No. 1, pp. 83-93, 2003.

[4] J. J. Aucounturier, and F. Pachet "Improving timbre similarity: How high is the sky?", *Journal Negative Results in Speech and Audio Sciences*, Vol. 1, No. 1, 2004.

[5] S. Bacon and D. Grantham, "Modulation masking: Effects of modulation frequency, depth, and phase", *Journal Acoustical Society of America*, Vol. 85, pp. 2575-2580, 1989.

[6] J. G. A. Barbedo and A. Lopes, "Automatic genre classification of musical signals", *EURASIP Journal Advances in Signal Processing*, Vol. 2007, Article ID 64960, 2007.

[7] D Bella, "Differentiation of classical music requires little learning but rhythm", *Cognition*, Vol. 96, No. 2 , pp 65-78, 2005.

[8] Ε. Μπενέτος. *Αναγνώριση Μουσικού Είδους με Τανυστές*. Μεταπτυχιακή Διατριβή, Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης, Τμήμα Πληροφορικής, 2007.

[9] Benetos, E. and Kotropoulos C. "A tensor-based approach for automatic music genre classification", in Proc. *2008 European Signal Processing Conference*, Lausanne, Switzerland, 2008.

[10] E. Benetos, M. Kotti, and C. Kotropoulos, "Applying supervised classifiers based on non-negative matrix factorization to musical instrument classification," in Proc. *IEEE Int. Conf. Multimedia & Expo*, July 2006.

[11] E. Benetos, M. Kotti, and C. Kotropoulos, "Large scale musical instrument identification," in Proc. *4th Sound and Music Computing Conf.*, July 2007.

[12] J. Bergstra, N. Casagrande, D. Erhan, D. Eck, and B. Kegl, "Aggregate features and AdaBoost for music classification", *Machine Learning*, Vol. 65, No. 2-3, pp. 473-484, 2006.

[13] D. Bertsekas, *Nonlinear Programming*, (2nd ed.) Belmont, MA: Athena Scientific,1999

[14] G. Birkhoff and S. Mac Lane, "*A Survey of Modern Algebra*", New York: Macmillan Publishing Co., 1977.

[15] A. I. Borisenko and I. E. Taparov, *Vector and Tensor Analysis with Applications*, New York: Dover Publications Inc., 1968.

[16] B.E. Boser, I.M. Guyon, and V. Vapnik, "A training algorithm for optimal margin classifiers", in Proc. *5th Annual Workshop Computational Learning Theory*, pp. 144-152, 1992.

[17] C. Boutsidis, E. Gallopoulos, P. Zhang, and R. Plemmons, "PALSIR: A new approach to nonnegative tensor factorization", Poster presented at *Workshop Algorithms for Modern Massive Data Sets*, June 2006.

[18] L. M. Bregman, "The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex programming", *USSR Computational Mathematics and Mathematical Physics*, Vol. 7, pp. 200-217, 1967.

[19] R. Bro, "PARAFAC: tutorial and applications", *Chemometrics and Intelligent Laboratory Systems*, Vol. 38, pp. 149-171, 1997.

[20] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, Vol. 2, pp. 121-167, 1998.

[21] J.J. Burred and A. Lerch, "A hierarchical approach to automatic musical genre classification", in Proc. *6th Int. Conf. Digital Audio Effects (DAFx)*, September 2003.

[22] D. Cai, X. He, J. Han, and H. J. Zhang, "Orthogonal laplacianfaces for face recognition", *IEEE Transactions Image Processing* , Vol. 15, no. 11, pp. 3608-3614, 2006.

[23] Z. Cataltepe, Y. Yaslan, and A. Sonmez, "Music genre classification using MIDI and audio features," *EURASIP Journal Advances in Signal Processing*, Vol. 2007, Article ID 36409, 2007.

[24] A. Chase, "Music discriminations by carp", *Animal Learning and Behavior*, Vol. 29, No. 4, pp. 336-353, 2001.

[25] T. Chi, Y. Gao, M. Guyton, P. Ru, and S. Shamma, "Spectrotemporal modulation transfer functions and speech intelligibility", *Journal Acoustical Society of America*, Vol. 106, No. 5, pp. 2719-2732, 1999.

[26] T. Chi, P. Ru, and S. Shamma, "Multiresolution spectrotemporal analysis of complex sounds", *Journal Acoustical Society of America*, Vol. 118, pp. 887-906, 2005.

[27] A. Cichocki, R. Zdunek, S. Choi, R. Plemmons, and S. Amari, "Non-negative tensor factorization using alpha and beta divergences", in Proc. *2007 IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, April 2007.

[28] Y. C. Cho, S. Choi, "Nonnegative features of spectro-temporal sounds for classification", *Pattern Recognition Letters*, Vol. 26, pp. 1327 - 1336, 2005.

[29] M. Chu, F. Diele, R. Plemmons, and S. Ragni, "Optimality, computation and interpretation of nonnegative matrix factorizations", *Preprint*, 2005, Available online at http://www4.ncsu.edu/ mtchu/Research/Papers/nnmf.ps

[30] A. Cichocki, R. Zdunek, S. Choi, R. Plemmons, and S. Amari, "Novel multi-layer nonnegative tensor factorization with sparsity constraints", in Proc. *2007 IEEE Int. Conf. Adaptive and Natural Computing Algorithms*, April 2007.

[31] A. Cichocki, R. Zdunek, S. Choi, R. Plemmons, and S. Amari, "Non-negative tensor factorization using alpha and beta divergences", in Proc. *2007 IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Honolulu, Hawai, 2007.

[32] A. Cichocki, R. Zdunek, and S. Amari, "Nonnegative Matrix and Tensor Factorization", *IEEE Signal Processing Magazine*, January 2008, pp. 142-145, 2008.

[33] P. Cook, *Music, Cognition, and Computerized Sound: An Introduction to Psychoacoustics*, The MIT Press, 1999.

[34] C. Cortes and V. Vapnik, "Support-Vector Networks," *Machine Learning*, Vol. 20, pp. 273-297, 1995.

[35] D. Depireux, J. Simon, D. Klein, and S. Shamma, "Spectrotemporal response field characterization with dynamic ripples in ferret primary auditory cortex", *Journal Neurophysiology*, Vol. 85, No. 3, pp. 1220-1234, 2001.

[36] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman, "Indexing by latent semantic analysis", *J. American Society for Information Science*, Vol. 41, No. 6, pp. 391-407, 1990.

[37] D. Donoho, and V. Stodden, "When does non-negative matrix factorization give a correct decomposition into parts?", in S. Thrun, L. Saul, and B. Schölkopf, eds., *Advances in Neural Information Processing Systems 16*, Cambridge: MIT Press, 2004.

[38] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed., New York: John Wiley & Sons, November 2000.

[39] H. Dudley, "Remaking speech", *Journal Acoustical Society of America*, Vol. 11, No. 2, pp. 169-177, 1939.

[40] J. Duchene, and S. Leclercq, "An optimal transformation for discriminant and principal component analysis", *IEEE Transactions Pattern Analysis and Machine Intelligence*, Vol. 10, No. 6, pp. 978-983, 1988.

[41] M. Elhilali, T. Chi, and S. Shamma, "A spectro-temporalmodulation index (STMI) for assessment of speech intelligibility", *Speech Communication*, to appear.

[42] S. Ewert and T. Dau, "Characterizing frequency selectivity for envelope fluctuations," *Journal Acoustical Society of America*, Vol. 108, pp. 1181-1196, 2000.

[43] D. Fitzgerald, M. Cranitch and E. Coyle, "Non-negative tensor factorization for sound source separation", in Proc. *Irish Signals and Systems Conference*, 2005.

[44] V. Franc, V. Hlavac, and M. Navara. "Sequential coordinate-wise algorithm for the non-negative least squares problem", in A. Gagalowicz and W. Philips, eds., *CAIP 2005: Computer Analysis of Images and Patterns*, Vol. 1, pp 407-414, Springer-Verlag, September 2005.

[45] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed., Baltimore MD: Johns Hopkins University Press, 1996.

[46] E. Gómez, A. Klapuri, and B. Meudic, "Melody description and extraction in the context of music content processing", *Journal New Music Research*, Vol. 32 No. 1, 2003.

[47] I. Guyon, J. Makhoul, R. Schwartz, and V. Vapnik, "What size test set gives good error rate estimates?", *IEEE Transactions Pattern Analysis and Machine Intelligence*, Vol. 20, No. 1, pp. 52-64, January 1998.

[48] F. Gouyon, S. Dixon, E. Pampalk, and G. Widmer, "Evaluating rhythmic descriptors for musical genre classification", in Proc. *AES 25th Int. Conf.*, pp 196-204, June 2004.

[49] W. H. Greub, *Multilinear Algebra*, New York: Springer-Verlag, 1967.

[50] L. Grippo, L. and M. Sciandrone, "On the convergence of the block nonlinear Gauss-Seidel method under convex constraints", *Operations Research Letters*, Vol. 26, pp. 127-136, 2000.

[51] T. Hazan, S. Polak, and A. Shashua, "Sparse image coding using a 3D non-negative tensor factorization", in Proc. *10th IEEE Int. Conf. Computer Vision*, Vol. 1, pp. 50-57, October 2005.

[52] F. van der Hedjen, R. P. W. Duin, D. de Ridder, and D. M. J. Tax, *Classification, Parameter Estimation and State Estimation*, London UK: Wiley, 2004.

[53] M. Heiler and C. Schnörr, "Controlling sparseness in non-negative tensor factorization", in Proc. *9th European Conf. Computer Vision*, Vol. 1, pp. 56-67, May 2006.

[54] H. Hermansky, "Should Recognizers Have Ears", *Special Issue on Robust Speech Recognition*, Vol. 25 , No. 1-3, pp. 3 - 27, 1998.

[55] T. Hofmann, "Probabilistic latent semantic analysis," in Proc. *Fifteenth Conf. Uncertainity in Artificial Intelligence*, pp. 289-296, July 1999.

[56] A. Holzapfel, and Y. Stylianou, "Musical genre classification using nonnegative matrix factorization-based features", *IEEE Transactions Audio, Speech, and Language Processing*, Vol. 16, No. 2, pp. 424-434, 2008.

[57] H. Homburg,I. Mierswa,B. Moller, K. Morik, M. Wurst, "Benchmark Dataset for Audio Classification and Clustering," in Proc. *Sixth Int. Symp. on Music Information Retrieval 2005*, pp. 528-531, London, 2005

[58] T. Houtgast, "Frequency selectivity in amplitude-modulation detection", *Journal Acoustical Society of America*, Vol. 85, pp. 1676-1680, 1989.

[59] P. O. Hoyer, "Non-negative matrix factorization with sparsness constraints", *Journal Machine Learning Research*, Vol. 5, pp. 1457-1469, 2004.

[60] C. Hsu, C. C. Chang, and C. J. Lin, *A Practical Guide to Support Vector Classification.* Technical Report, Department of Computer Science, National Taiwan University, 2003.

[61] C. Hu, B. Zhang, S. Yan, Q. Yang, J. Yan, Z. Chen, and W. Ma, "Mining ratio rules via principal sparse non-negative matrix factorization," in Proc. *2004 IEEE Int. Conf. Data Mining*, 2004.

[62] A. Hyvärinen and E. Oja, Independent component analysis: Algorithms and applications, *Neural Networks*, Vol. 13, pp. 411-430, 2000.

[63] B. Johanssona, T. Elfvingb, V. Kozlovc, Y. Censord, P. E. Forssona, and G. Granlund, "The application of an oblique-projected Landweber method to a model of supervised learning", *Mathematical and Computer Modelling*, Vol, 43, No. 7-8, pp. 892-909, April 2006.

[64] A. Kapteyn, H. Neudecker, and T. Wansbeek, "An approach to n-mode components analysis," *Psychometrica*, Vol. 51, No. 2, pp. 269-275, June 1986.

[65] D. C. Kay, *Theory and Problems of Tensor Calculus*. New York: McGraw-Hill, 1988.

[66] T. G. Kolda, "Orthogonal tensor decompositions," *SIAM Journal Matrix Analysis Applications*, Vol. 23, No. 1, pp. 243-255, 2001.

[67] I. Kotsia , S. Zafeiriou and I. Pitas , "A novel discriminant non-negative matrix factorization algorithm with applications to facial image characterization problems",*IEEE Transactions Forensics and Security*, Vol. 2, No. 3, part 2, pp. 588-595, 2007.

[68] N. Kowalski, D. Depireux, and S. Shamma, "Analysis of dynamic spectra in ferret primary auditory cortex: I. Characteristics of single unit responses to moving ripple spectra" *Journal Neurophysiology*, Vol. 76, No. 5, pp. 3503-3523, 1996.

[69] P. M. Kroonenberg and J. De Leeuw, "Principal component analysis of three-mode data by means of alternating least squares algorithms", *Psychometrika*, Vol. 45, No.1, March 1980.

[70] G. Langner, "Periodicity coding in the auditory system", *Hearing Research*, Vol. 60, No. 2, pp. 115-142, 1992.

[71] G. Langner, M. Sams, P. Heil, and H. Schulze, "Frequency and periodicity are represented in orthogonal maps in the human auditory cortex: evidence from magnetoencephalography", *Journal of Comparative Physiology*, Vol. 181, No. 6, pp. 665-676, 1997.

[72] L. De Lathauwer, "*Signal Processing Based on Multilinear Algebra*", Ph.D. thesis, K.U. Leuven, E.E. Dept.-ESAT, Belgium, 1997.

[73] L. De Lathauwer, B. De Moor, and J. Vandewalle "A multilinear singular value decomposition", *SIAM J. Matrix Analysis Applications*, Vol. 21, No. 4, pp. 1253-1278, April 2000.

[74] D. D. Lee, and H. S. Seung , "Algorithms for non-negative matrix factorization", *Advances in Neural Information Processing Systems*, No. 13, pp. 556-562, 2001.

[75] J. H. Lee, and J. S. Downie, "Survey of music information needs, uses and seeking behaviours: Preliminary findings", in Proc. *Fifth Int. Symp. on Music Information Retrieval*, Barcelona, Spain, 2004.

[76] T. Li, M. Ogihara, and Q. Li, "A comparative study on content-based music genre classification", in Proc. *26th Annual International ACM SIGIR Conf. on Research and Development in Informaion Retrieval*, Toronto, Canada, 2003

[77] C.-J. Lin, "Projected gradient methods for nonnegative matrix factorization", *Neural Computation*, Vol. 19 , No. 10 , pp 2756-2779, 2007.

[78] T. Lidy, and A. Rauber, "Evaluation of feature extractors and psycho-acoustic transformations for music genre classification", in Proc. *Sixth Int. Symp. on Music Information Retrieval*, London, UK, 2005.

[79] T. Lidy, A. Rauber, A., Pertusa, and J. Inesta, "Combining audio and symbolic descriptors for music classification from audio", *Music Information Retieval Information Exchange (MIREX)*, 2007.

[80] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "MPCA: Multilinear Principal Component Analysis of Tensor Objects", *IEEE Transactions Neural Networks*, Vol. 19, No. 1, pp 18-39, 2008.

[81] S. Z. Li, X. Hou, H. Zhang, and Q. Cheng, "Learning spatially localized, parts-based representation," in Proc. *IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1-6, 2001.

[82] L. Lim, "Optimal solutions to non-negative PARAFAC/Multilinear NMF always exist", in Proc. *Workshop Tensor Decompositions and Applications*, August-September 2005.

[83] R.F. Lyon, "A computational model of filtering, detection and compression in the cochlear", in Proc. *IEEE Int. Conf. Acoust., Speech, & Sig. Proc.*, Paris, 1982, pp. 1282-1285.

[84] M. Mandel, and D. Ellis, "Song-level features and support vector machines for music classification", in Proc. *Sixth Int. Symp. on Music Information Retrieval*, London, UK, 2005.

[85] Mandel, M. and Ellis, D. "Labrosas: audio music similarity and classification submissions", *Music Information Retieval Information Exchange (MIREX)*, 2007.

[86] C. McKay, and I. Fujinaga, "Musical genre classification: Is it worth pursuing and how can in be improved?", in Proc. *Seventh Int. Symp. on Music Information Retrieval*, Victoria, Canada, 2006.

[87] M.M. Merzenich, P.L. Knight, and G.L. Roth, "Representation of cochlea within the primary auditory cortex in cat", *Journal Neurophysiology*, Vol. 28, pp. 231-249, No. 2, 1975.

[88] N. Mesgarani, M. Slaney, S. A. and Shamma S., "Discrimination of speech from non-speech based on multiscale spectro-temporal modulations", *IEEE Transactions Audio, Speech and Language Processing*, Vol. 14, pp. 920-930, 2006.

[89] M. Mandel and D. Ellis, "Song-level features and support vector machines for music classification", in Proc. *Sixth Int. Symp. on Music Information Retrieval*, pp. 594-599, September 2005.

[90] A. Meng, P. Ahrendt, and J. Larsen, "Improving music genre classification by short-time feature integration", in Proc. *6th Int. Symp. Music Information Retrieval*, pp. 604-609, September 2005.

[91] A. Meng, "General Purpose Multimedia Dataset - GarageBand 2008", Technical University of Denmark, 2008.

[92] P. Merwe, *Origins of the Popular Style - The Antecedents of Twentieth-Century Popular Music*. Oxford University Press, 1989.

[93] A.Moller, "Unit responses of the rat cochlear nucleus to tones of rapidly varying frequency and amplitude", *Acta Physiol. Scan.*, Vol. 81, pp. 540-556, 1971.

[94] R. Munkong, and B.-H. Juang, "Auditory perception and cognition", *IEEE Signal Processing Magazine*, Vol. 25, No. 3, pp. 98-117, 2008.

[95] F. Pachet and D. Cazaly, "A taxonomy of musical genres", in Proc. *Content-Based Multimedia Information Access Conf.*, April 2000.

[96] F. Pachet, J.J. Aucouturier, A. La Burthe, A. Zils, and A. Beurive, "The CUIDADO music browser: an end-to-end electronic music distribution system", *Multimedia Tools and Applications, Special Issue on the CBMI 2003 Conf.*, 2004.

[97] E. Pampalk, A. Flexer, and G. Widmer, "Improvements of audio based music similarity and genre classification", in Proc. *Sixth Int. Symp. Music Information Retrieval*, pp. 628-633, 2005.

[98] A. Papoulis, *Probability, Random Variables and Stochastic Processes*, 2nd ed., New York: McGraw-Hill, 1984.

[99] I. Panagakis, E. Bennetos, C. Kotropoulos, "Music genre classification: A multilinear approach", in Proc. *Ninth Int. Symp. on Music Information Retrieval*, Philadelphia, PA USA , 2008

[100] G. Peeters, "A large set of audio features for sound description (similarity and classification) in the CUIDADO project", *CUIDADO I.S.T. Project Report*, 2004.

[101] D. Perrott and R.O. Gjerdingen, "Scanning the dial: An exploration of factors in the identification of musical style", *Dept. Music, Northwestern University, Illinois, Res. Notes*, 1999.

[102] J. G. Proakis, C. M. Rader, F. Ling, C. L. Nikias, M. Moonen, and I. K. Proudler, *Algorithms for Statistical Signal Processing*, Upper Saddle River, New Jersey: Prentice Hall, 2002.

[103] S. Ravindran, K. Schlemmer and D. Anderson, " A Physiologically Inspired Method for Audio Classification," *EURASIP Journal Applied Signal Processing*, Vol. 2005, No. 9, pp. 1374-1381, 2005.

[104] A. Rauber, E. Pampalk, and D. Merkl, "Using psycho-acoustic models and selforganizing maps to create a hierarchical structuring of music by sound similarity," in Proc. *3rd Int. Conf. Music Information Retrieval*, October 2002.

[105] F. Rousseaux and A. Bonardi, "Reconcile art and culture on the Web: Lessen the importance of instantiation so creation can better be fiction," in Proc. *1st Int. Workshop Philosophy Informatics*, April 2004.

[106] R. Rifkin, and N. Mesgarani, "Discriminating Speech and Non-Speech with Regularized Least Squares", in Proc. *Ninth International Conference on Spoken Language Processing*, Pittsburgh, USA, 2006.

[107] A. Shashua, and T. Hazan, "Non-negative tensor factorization with applications to statistics and computer vision", in Proc. *22nd International Conference on Machine Learning*, 2005.

[108] N. Sidiropoulos, and R. Bro, "On the uniqueness of multilinear decompositions", *Journal Chemometrics*, Vol. 14, pp. 229-239, 2000.

[109] N. C. Singh, and F. E. Theunissen, "Modulation spectra of natural sounds and ethological theories of auditory processing", *Journal Acoustical Society of America*, Vol. 114, pp. 3394-3411, 2003.

[110] S. Sundaram, S. Narayanan, "Discriminating two types of noise sources using cortical representation and dimension reduction technique" in Proc. *2007 IEEE International Conference on Acoustics, Speech, and Signal Processing*, Honolulu, Hawai, 2007.

[111] N. Scaringella and G. Zoia, "On the modeling of time information for automatic genre recognition systems in audio signals," in Proc. *6th Int. Symp. Music Information Retrieval*, pp. 666-671, September 2005.

[112] C. Schreiner and J.Urbas, "Representation of amplitude modulation in the auditory cortex of the cat I. The anterior auditory field (AAF)", *Hearing Research*, Vol. 21, pp. 227-241, 1986.

[113] H. Schulze and G. Langner, "Periodicity coding in the primary auditory cortex of the Mongolian gerbil (Meriones unguiculatus): two different coding strategies for pitch and rhythm", *Journal Comparative Physiology A*, Vol. 181, No. 6, pp. 651-663, 1997.

[114] N. Scaringella, G. Zoia, and D. Mlynek, "Automatic genre classification of music content: A survey", *IEEE Signal Processing Mag.*, Vol. 23, No. 2, pp. 133-141, March 2006.

[115] S.A. Shamma, J.W. Fleshman, P.W. Wiser, and H. Versnel, "Organization of response areas in ferret primary auditory cortex", *Journal Neurophysiology*, Vol. 69, No. 2, pp. 367-383, 1993.

[116] S. Shamma, "Auditory cortical representation of complex acoustic spectra as inferred from the ripple analysis method", Network: Computation in Neural Systems, Vol. 7, No. 3, pp. 439-476, 1996.

[117] A. Shashua and T. Hazan, "Non-negative tensor factorization with applications to statistics and computer vision," in Proc. *22nd Int. Conf. Machine Learning*, pp. 792-799, August 2005.

[118] X. Shao, C. Xu, and M. Kankanhalli, "Unsupervised classification of musical genre using hidden Markov model", in Proc. *2004 IEEE Int. Conf. Multimedia & Expo*, pp. 2023-2026, 2004.

[119] S. Sheft and W. Yost, "Temporal integration in amplitude modulation detection", *Journal Acoustical Society of America*, Vol. 88, pp. 796-805, 1990.

[120] M. Slaney and R. Lyon, "The importance of time - a temporal representation of sound", In *M. Cooke, S. Beet, and M. Crawford, eds., Visual Representations of Speech Signals.* John Wiley and Sons, New York, 1993.

[121] H. Soltau, T. Schultz, M. Westphal, and A. Waibel, "Recognition of music types", in Proc. *IEEE Int. Conf. Acoustics, Speech Signal Processing*, Vol. II, pp. 1137-1140, 1998.

[122] Χ. Σπυρίδης, Μια Εισαγωγή στη Φυσική της Μουσικής, Εκδ. Ζήτη, 1988

[123] N. Suga, "Analysis of information-bearing elements in complex sounds by auditory neurons of bats", Audiology, Vol. 11, pp. 58-72, 1972.

[124] S. Sukittanon, L.E. Atlas, J.W. Pitton, "Modulation-scale analysis for content identification", *IEEE Transactions Acoustics, Speech, and Signal Processing,* Vol. 52, No. 10, pp. 3023-3035, 2004.

[125] S. Sukittanon, L. Atlas, J. Pitton, and K. Filali, "Improved Modulation Spectrum Through Multi-scale Modulation Frequency Decomposition", In Proc. *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Philadelphia, Pennsylvania, 2005.

[126] G. Tzanetakis, and P. Cook, "Musical genre classification of audio signal", *IEEE Transactions Speech and Audio Processing*, Vol. 10, No. 3, pp. 293-302, July 2002.

[127] L. R. Tucker, "The extension of fator analysis to three-dinensional matrices," in H. Gulliksen, N. Frederiksen, *Contributions to Mathematical Psycology*, Holt, Rinehart & Winston, N.Y., pp.109-127, 1964.

[128] V. Vapnik, *Statistical Learning Theory*. J. Wiley, New York, 1998.

[129] K. West and S. Cox, "Finding an optimal segmentation for audio genre classification", in Proc. *Sixth Int. Symp. Music Information Retrieval*, pp. 680-685, September 2005.

[130] K. Wang, and S. A. Shamma, "Spectral shape analysis in the central auditory system", *IEEE Transactions on Speech and Audio Processing*, Vol. 3, pp. 382-396, 1995.

[131] M. Welling, and M. Weber, "Positive tensor factorization", *Pattern Recognition Letters*, Vol. 22, No. 12, pp. 1255-1261, 2001.

[132] M. Wohlmayr, M. Markaki, and Y. Stylianou, "Speech - Nonspeech Discrimination Based on Speech-Relevant Spectrogram Modulations" *Proceedings of the European Signal Processing Conference*, Poznan, Poland, 2007.

[133] S. Woolley, T. Fremouw, A. Hsu, and F. Theunissen, "Tuning for spectro-temporal modulations as a mechanism for auditory discrimination of natural sounds", *Nature Neuroscience*, Vol. 8, pp. 1371-1379, 2005.

[134] E. Zwicker and E. Terhardt, "Analytical expressions for critical-band rate and critical bandwidth as a function of frequency", *Journal Acoustical Society of America*, Vol. 68, No. 5, pp. 1523-1525, June 1980.

[135] Εγκυκλοπαίδεια Grand Larousse.