

Πτυχιακή Εργασία

Αναγνώριση συναισθήματος σε μουσικά κομμάτια

Βάσωφ Αθανάσιος

2014



Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών
Υπολογιστών
Πολυτεχνική Σχολή του Πανεπιστημίου Θεσσαλίας



ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ
ΒΙΒΛΙΟΘΗΚΗ & ΚΕΝΤΡΟ ΠΛΗΡΟΦΟΡΗΣΗΣ
ΕΙΔΙΚΗ ΣΥΛΛΟΓΗ «ΓΚΡΙΖΑ ΒΙΒΛΙΟΓΡΑΦΙΑ»

Αριθ. Εισ.: 12420/1
Ημερ. Εισ.: 04-04-2014
Δωρεά: Συγγραφέα
Ταξιθετικός Κωδικός: ΠΤ – ΗΜΜΥ
2014
ΒΑΣ

Επιβλέπων Καθηγητής

Ποταμιάνος Γεράσιμος, Αναπληρωτής Καθηγητής

Δεύτερο μέλος Επιτροπής

Κατσαβουνίδης Ιωάννης, Αναπληρωτής Καθηγητής

Contents

1	Εισαγωγή	9
1.1	Σημασία της Αναγνώρισης Συναισθήματος Μουσικής	9
1.2	Προβλήματα στην Αναγνώριση Συναισθήματος Μουσικής	10
1.3	Συναισθηματικά Μοντέλα στον Τομέα της Ψυχολογίας	11
2	Εξαγωγή Χαρακτηριστικών	13
2.1	Εισαγωγή	13
2.2	Mel Frequency Cepstral Coefficients (MFCC)	14
2.3	Δυναμικά Χαρακτηριστικά	14
2.3.1	Ενέργεια Ρίζας Μέσου Τετραγώνου	15
2.3.2	Χρόνος Επίθεσης και Κλίση Επίθεσης	15
2.3.3	Ποσοστό Χαμηλής Ενέργειας	17
2.4	Ρυθμικά Χαρακτηριστικά	18
2.4.1	Καμπύλη Ανίχνευσης Γεγονότων	18
2.4.2	Αυτοσυσχέτιση	19
2.4.3	Υπολογισμός Περιοδικότητας	19
2.4.4	Τέμπο και Αλλαγή Τέμπο	20
2.4.5	Μετρικό κεντροειδές και Μετρική Ισχύς	20
2.4.6	Καθαρότητα Παλμών	21
2.4.7	Πυκνότητα Γεγονότων	22
2.5	Φασματικά Χαρακτηριστικά	22
2.5.1	Ρυθμός Μηδενικής Διασταύρωσης	23
2.5.2	Φασματική Κύλιση	23
2.5.3	Φασματικό Κεντροειδές	24
2.5.4	Φασματική Έκταση	24
2.5.5	Φασματική λοξότητα	24
2.5.6	Εντροπία	24
2.5.7	Φασματική Ροή	25
2.5.8	Φασματική ομαλότητα	25
2.5.9	Παρατυπία	25
2.6	Αρμονικά Χαρακτηριστικά	26
2.6.1	Χρωμόγραμμα	26
2.6.2	Ισχύς Τονικότητας, Καθαρότητα Τονικότητας και Τρόπος	27
2.6.3	Τονικό Κεντροειδές και Hcdf	28
2.6.4	Τραχύτητα	31
2.6.5	Δυσαρμονία	32

3 Ταξινομητές	35
3.1 Εισαγωγή	35
3.2 Κατηγορίες Ταξινόμησης σε Συστήματα Αναγνώρισης Μουσικού Συναισθήματος	35
3.3 Μοντέλο Μείγματος Γκαουσιανών Κατανομών	36
3.4 Μηχανές Διανυσματικής Στήριξης	37
4 Επιλογή Χαρακτηριστικών	39
4.1 Εισαγωγή	39
4.2 Μέθοδοι Επιλογής Χαρακτηριστικών	39
4.3 Μέθοδοι Σταδιακής Επιλογής (Stepwise Selection)	40
4.4 Μέθοδος Αναρρίχησης Λόφου Τυχαίας Μετάλλαξης (Random Mutation Hill Climbing)	40
5 Συλλογή της Βάσης Δεδομένων και Προεπεξεργασία	43
5.1 Εισαγωγή	43
5.2 Η Συλλογή της Βάσης Δεδομένων	43
5.3 Προεπεξεργασία των Δεδομένων (Data Pre-Processing)	45
6 Πειράματική Διαδικασία και Αποτελέσματα	47
6.1 Πείραμα 1: Εκπαίδευση GMM με Χαρακτηριστικά MFCC	47
6.1.1 Εξαγωγή των MFCC	47
6.1.2 Εκπαίδευση των GMM	47
6.1.3 Αποτελέσματα	48
6.2 Πείραμα 2: Εκπαίδευση GMM με το Σύνολο των Χαρακτηριστικών	49
6.2.1 Εξαγωγή των Χαρακτηριστικών	49
6.2.2 Εκπαίδευση των GMM	49
6.2.3 Συμπεράσματα	49
6.3 Πείραμα 3: Βελτίωση των αποτελεσμάτων μέσω γραμμικού συνδιασμού των Γκαουσιανών	51
6.3.1 Πειραματική Διαδικασία	51
6.3.2 Συμπεράσματα	51
6.4 Πείραμα 4: Επιλογή Χαρακτηριστικών για Εκπαίδευση με GMM	52
6.4.1 Οπισθοδρομική Απαλοιφή Χαρακτηριστικών	52
6.4.2 Αναρρίχηση Λόφου Τυχαίας Μετάλλαξης	52
6.4.3 Συμπεράσματα	53
6.5 Πείραμα 5: Εκπαίδευση με SVM	54
6.5.1 Υλοποίηση του SVM	54
6.5.2 Πειραματική Διαδικασία	55
6.5.3 Συμπεράσματα	55
6.6 Πείραμα 6: Επιλογή Χαρακτηριστικών για Εκπαίδευση με SVM	55
6.6.1 Εμπρόσθια Επιλογή Χαρακτηριστικών	55
6.6.2 Αναρρίχηση Λόφου Τυχαίας Μετάλλαξης	56
6.6.3 Συμπεράσματα	56

7	Συμπεράσματα	59
7.1	Συνολικά Συμπεράσματα των Πειραμάτων	59
7.2	Συμβολή της Διπλωματικής Εργασίας	59
7.3	Μελλοντικές Ερευνητικές Κατευθύνσεις	60

	Bibliography	65
--	---------------------	-----------

List of Tables

1.1	Κατανομή ετικετών χρηστών (από [23])	10
2.1	Διανυσμα Χαρακτηριστικών	33
5.1	Κατηγορίες που προέκυψαν και παραδείγματα ετικετών	46
5.2	Συνολικός αριθμός δειγμάτων mp3 ανά κατηγορία	46
6.1	Αποτελέσματα Εκπαίδευσης GMM με χαρακτηριστικά MFCC (όπου μ ο μέσος όρος, σ η τυπική απόκλιση και Δ η παράγωγος). Με έντονη γραμμοτοσειρά εμφανίζεται η μέγιστη επιτυχία ανά κατηγορία.	48
6.2	Αποτελέσματα Εκπαίδευσης GMM μία Γκαουσιανής κατανομής ανά κατηγορία χαρακτηριστικών. Με έντονη γραμμοτοσειρά εμφανίζεται η μέγιστη επιτυχία ανά κατηγορία.	50
6.3	Βέλτιστα βάρη για συνδιασμό Γκαουσιανών και αποτελέσματα ταξινόμησης.	52
6.4	Βέλτιστο υποσύνολο και επιτυχία Οπισθοδρομικής Απαλοιφής για GMM.	53
6.5	Βέλτιστο υποσύνολο και επιτυχία RMHC για GMM.	54
6.6	Αποτελέσματα Εκπαίδευσης SVM.	55
6.7	Βέλτιστο υποσύνολο και επιτυχία FFS για SVM.	56
6.8	Βέλτιστο υποσύνολο και επιτυχία RMHC για SVM.	57

List of Figures

1.1	Μοντέλο συναισθημάτων του Henver (από [18])	12
1.2	Μοντέλο συναισθημάτων του Russell (από [33])	12
2.1	Εξαγωγή Χαρακτηριστικών MFCC	14
2.2	Υπολογισμός χρόνου επίθεσης (από [26])	16
2.3	Υπολογισμός κλίσης επίθεσης (από [26])	16
2.4	Φάκελος κυματομορφής του κομματιού ragtime (από [26])	17
2.5	Παράδειγμα υπολογισμού Ποσοστού Χαμηλής Ενέργειας (από [26])	18
2.6	Αυτοσυσχέτιση ενός στιγμιότυπου τρομπέτας (από [26])	19
2.7	Τέμπο και Αλλαγή Τέμπο (από [26])	21
2.8	Μετρικό κεντροειδές και Μετρική ισχύς (από [26])	22
2.9	Το μέγιστο (μαύρος κύκλος) της συνάρτησης αυτοσυσχέτισης μας δίνει την καθαρότητα του ρυθμού (από [25])	22
2.10	Παράδειγμα roll-off 85% (από [26])	23
2.11	Φάσμα και φασματική ροή (από [26])	25
2.12	Παράδειγμα χρωμογράμματος (από [26])	27
2.13	Παράδειγμα διασυσχέτισης ισχύος κλειδιού: String Quartet Op. 30 I Moderato, from Arnold Schoenberg (από [16])	28
2.14	Παράδειγμα ισχύος τονικότητας, τονικότητας και καθαρότητας τονικότητας (από [26])	29
2.15	Παράδειγμα υπολογισμού τρόπου (από [26])	29
2.16	Tonnetz Harmonic Network (από [17])	30
2.17	Οπτικοποίηση του 6-D τονικού χώρου ως τρεις κύκλοι. Από αριστερά προς τα δεξιά: Πέμπτης, Τρίτης μικρής, Τρίτης μεγάλης. Το σημείο A είναι τονικό κεντροειδές της συγχορδίας Λα Ματζόρε (ημιτόνια 9,1,4) (από [17])	30
2.18	Μοντέλο τραχύτητας Plomp Levelt (από [26])	31

Περίληψη

Η παρούσα διπλωματική έχει σαν αντικείμενο τη μελέτη της αυτόματης αναγνώρισης συναισθήματος/διάθεσης σε σήματα μουσικής. Εφαρμογές αυτής της μελέτης είναι η αυτόματη κατηγοριοποίηση αρχείων μουσικής και η βελτίωση εφαρμογών ανάκτησης μουσικής πληροφορίας. Αρχικά, αφού πρώτα εισάγουμε τον αναγνώστη στο αντικείμενο, γίνεται περιγραφή των χαρακτηριστικών που, σύμφωνα με την βιβλιογραφία, προσφέρουν πληροφορία μουσικής αντίληψης. Επίσης, γίνεται ανάλυση του τρόπου εξαγωγής αυτών από ακουστικά σήματα. Στη συνέχεια, εστιάζουμε στην εξέταση των αλγόριθμων ταξινόμησης που χρησιμοποιούνται συνήθως για κατηγοριοποίηση μουσικών δειγμάτων και περιγράφουμε συνοπτικά τους αλγόριθμους επιλογής χαρακτηριστικών που χρησιμοποιήσαμε στα πειράματά μας. Σημαντικό τμήμα της διπλωματικής εργασίας είναι η περιγραφή της σύνθεσης δικής μας βάσης δεδομένων, η οποία, από όσο γνωρίζουμε, είναι η μεγαλύτερη βάση δεδομένων μουσικής-συναισθήματος στην βιβλιογραφία. Τέλος, γίνεται περιγραφή της πειραματικής διαδικασίας και παράθεση των αποτελεσμάτων. Τα αποτελέσματά αυτά είναι πολύ ικανοποιητικά, με επιτυχία της τάξης του 60% για τέσσερις κλάσεις συναισθήματος, και είναι συγκρίσιμα με αντίστοιχες μελέτες πάνω στο αντικείμενο.

Abstract

This Diploma thesis focuses on automatic emotion/mood recognition in music signals. Some applications of this study include automatic music categorization and the improvement of music information retrieval applications. Initially, after an introduction on the subject, we describe the features useful in musical perception, according to the literature. We further analyse their extraction from acoustic signals. Continuing, we examine the classification algorithms usually employed in the taxonomy of music excerpts, and we concisely describe the feature selection algorithms used in our experiments. Furthermore, we describe the compilation of our own data base, which to our knowledge constitutes the most extensive music-emotion data base available in the literature. Finally, we describe the experimental procedure and present our results. These are quite satisfactory, reaching a success rate of over 60% for four emotion classes, and are comparable to the best results reported in the literature.

1 Εισαγωγή

Η σημαντικότερη λειτουργία της μουσικής είναι η δημιουργία συναισθημάτων. Κανείς δεν μπορεί να συνθέσει, να εκτελέσει ή να ακούσει μουσική χωρίς να επηρεαστεί συναισθηματικά. Η μουσική μπορεί να μας κάνει να δακρύσουμε και να μας συντροφεύσει στη λύπη και στη χαρά. Έρευνες, που μελετούν τη συμπεριφορά στη μουσική πληροφορία, έχουν αναγνωρίσει το συναίσθημα ως ένα σημαντικό κριτήριο στην αναζήτηση και κατηγοριοποίηση της μουσικής. Σε αυτό το κεφάλαιο, θα εξηγήσουμε τη σημασία της αυτόματης αναγνώρισης συναισθήματος μουσικής και θα αναφέρουμε τα προβλήματα που συναντώνται στον τομέα αυτό. Ακόμη, θα εξετάσουμε τα συναισθηματικά μοντέλα από την πλευρά της επιστήμης της ψυχολογίας.

1.1 Σημασία της Αναγνώρισης Συναισθήματος Μουσικής

Η μουσική αποτελεί ανέκαθεν σημαντικό κομμάτι της ανθρώπινης ζωής, πόσο μάλλον στην εποχή των ψηφιακών μέσων. Με την ανάπτυξη της υψηλής ταχύτητας διαδικτύου, τεράστιος όγκος μουσικής πληροφορίας είναι πλέον προσιτός από τους χρήστες. Ένα παράδειγμα είναι η ραγδαία ανάπτυξη μουσικών βιβλιοθηκών (YouTube, Last.FM, Spotify κ.ά). Σε αυτό έχει συντελέσει και η ανάπτυξη μορφών συμπίεσης ήχου, όπως το mp3 (MPEG-1 Audio Layer 3), οι οποίες προσφέρουν υψηλή ποιότητα σε μικρά (σε σχέση με την ταχύτητα του διαδικτύου) αρχεία. Συμβατικά, η διαχείριση των συλλογών μουσικής είναι βασισμένη στα μεταδεδομένα (metadata) των τραγουδιών, όπως το όνομα καλλιτέχνη ή τον τίτλο τραγουδιού. Όμως, καθώς οι απαιτήσεις των χρηστών για αναζήτηση και κατηγοριοποίηση αυξάνονται, αυτός ο τρόπος διαχείρισης σταδιακά αποδεικνύεται μη επαρκής.

Σύμφωνα με μια κοινωνιολογική μελέτη του 2007 για τη χρήση ετικετών (tags) κατηγοριοποίησης από τους χρήστες του Last.FM [3] οι ετικέτες συναισθήματος είναι ο τρίτος συχνότερος τρόπος κατηγοριοποίησης μουσικής που επιλέγεται (βλέπε Πίνακα 1.1). Από τότε, η ανάκτηση συναισθηματικής πληροφορίας μουσικής λαμβάνει αυξανόμενη προσοχή σε ακαδημαϊκό επίπεδο αλλά και στη βιομηχανία εφαρμογών. Για παράδειγμα, τα τελευταία χρόνια παρατηρούμε τη ραγδαία αύξηση δημοτικότητας ιστοσελίδων που προσφέρουν λίστες αναπαραγωγής διάθεσης/συναισθήματος όπως Stereomood [9], Musicoverly [5], το ελληνικό site Kasetophono [7] κ.ά.

Δημιουργώντας συστήματα αναγνώρισης συναισθήματος από μουσική, βελτιώνουμε τον τρόπο που οι άνθρωποι αλληλεπιδρούν με τους υπολογιστές. Είναι εφικτή η δημιουρ-

Table 1.1: Κατανομή ετικετών χρηστών (από [23])

Τύπος Ετικέτας	Συχνότητα Χρήσης	Παραδείγματα
Είδος	68%	Heavy metal, punk
Τοπικά	12%	French, Seattle, NYC
Διάθεση/Συναισθήμα	5%	Chill, party
Άποψη	4%	Love, favorite
Ενορχήστρωση	4%	Piano, female vocal
Στύλ	3%	Political, humor
Διάφορα	2%	Coldplay, composers
Προσωπικά	1%	Seen live, I own it
Οργανωτικά	1%	Check out

γία εφαρμογών που επιλέγουν μουσική υπόκρουση που σχετίζεται με την διάθεση των χρηστών. Αυτό μπορεί να γίνει μέσω συλλογής ψυχολογικών, προσωπικών ή άλλων στοιχείων όπως τα χαρακτηριστικά έκφρασης προσώπου. Για παράδειγμα, μπορούμε να εξοπλίσουμε φορητές συσκευές (π.χ. mp3 players, κινητά τηλέφωνα) με λειτουργικότητα αναγνώρισης συναισθήματος μουσικής ώστε να παίζουν πάντα μουσική σύμφωνα με τη διάθεση του χρήστη. Επίσης, μπορούμε να δημιουργήσουμε “έξυπνους” χώρους (π.χ. εστιατόρια, καφετέριες, οικίες) που επιλέγουν την αναπαραγωγή κατάλληλης μουσικής λίστας με βάση τη διάθεση των ανθρώπων που βρίσκονται μέσα.

1.2 Προβλήματα στην Αναγνώριση Συναισθήματος Μουσικής

Καθώς ο τομέας της αναγνώρισης συναισθήματος από μουσική είναι ακόμα σε ανάπτυξη, υπάρχουν διάφορα προβλήματα που πρέπει να αντιμετωπιστούν. Αναφέρουμε συνοπτικά τα σημαντικότερα από αυτά:

- Ασάφεια περιγραφής συναισθήματος: Τα ανθρώπινα συναισθήματα είναι αφηρημένες έννοιες. Αυτό συμβαίνει γιατί οι άνθρωποι παρόλο που συνήθως αναγνωρίζουν τα συναισθήματα τους, δυσκολεύονται να τα εκφράσουν με ακρίβεια και σταθερότητα. Για αυτό το λόγο, το να καθορίσουμε συγκεκριμένες κατηγορίες ταξινόμησης είναι ευαίσθητη και με σημαντικές δυσκολίες διαδικασία. Το ίδιο συμβαίνει και στη διαδικασία συλλογής βάσης δεδομένων καθώς δεν μπορούμε να είμαστε σίγουροι ότι όλα τα άτομα που συμμετείχαν στο σχολιασμό ενός κομματιού αντιλαμβάνονται το ίδιο συναίσθημα.
- Υποκειμενικότητα της συναισθηματικής αντίληψης: Η μουσική αντίληψη είναι σε τεράστιο βαθμό υποκειμενική και καθορίζεται από ποικίλους παράγοντες όπως: ο χαρακτήρας, η μουσική παιδεία, η ηλικία, το φύλο, οι κοινωνικές επιρροές κ.ά. Λόγω αυτής της υποκειμενικότητας είναι δύσκολο να συσχετίσουμε ένα μουσικό κομμάτι με το συναίσθημα που παράγει με γενικό και αντικειμενικό τρόπο.

- Σημασιολογικό χάσμα μεταξύ χαρακτηριστικών ήχου και ανθρώπινης αντίληψης: Υπάρχει τεράστια δυσκολία στο να κατανοήσουμε πώς λειτουργεί η συναισθηματική διέγερση ενός ανθρώπου. Παρά την προσπάθεια στους τομείς της ψυχολογίας και κοινωνιολογίας δε μπορούμε να εντοπίσουμε ποια ακριβώς είναι τα χαρακτηριστικά της μουσικής που επιδρούν συναισθηματικά σε κάθε άνθρωπο. Για αυτό το λόγο, η δημιουργία ενός συστήματος που προσομοιώνει την ανθρώπινη μουσική αντίληψη, δεδομένου χαρακτηριστικών μίας κυματομορφής, είναι περίπλοκη διαδικασία που παρουσιάζει σημαντικές δυσκολίες.

1.3 Συναισθηματικά Μοντέλα στον Τομέα της Ψυχολογίας

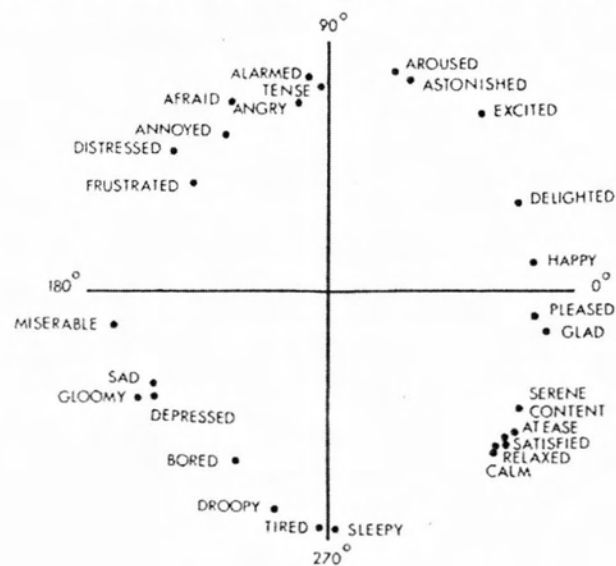
Η σχέση συναισθήματος και μουσικής έχει ερευνηθεί από ψυχολόγους για δεκαετίες. Σημαντικό πρόβλημα που τέθηκε είναι η κατασκευή ενός μοντέλου που καταγράφει όλα τα δυνατά συναισθήματα και τη συσχέτιση μεταξύ τους. Η μελέτη αυτών των μοντέλων είναι πολύ σημαντική για την κατασκευή ενός συστήματος αυτόματης αναγνώρισης συναισθημάτων. Τα περισσότερα από τα συναισθηματικά μοντέλα μπορούν να αντιστοιχηθούν σε μία εκ' των δύο παρακάτω προσεγγίσεων.

- Κατηγορηματική Προσέγγιση (Categorical Approach): Πρόκειται για μοντελοποίηση των συναισθημάτων σε ξεχωριστές κατηγορίες με σκοπό η κάθε κατηγορία να διαφοροποιείται από τις υπόλοιπες. Για να γίνει μια τέτοιου είδους μοντελοποίηση πρέπει να θεωρήσουμε ότι υπάρχει πεπερασμένος αριθμός συναισθημάτων, τα οποία μπορούν να ομαδοποιηθούν με τέτοιο τρόπο ώστε η κάθε κατηγορία να αντιστοιχεί σε ένα βασικό συναίσθημα. Το διασημότερο κατηγορηματικό μοντέλο αναπτύχθηκε από την K. Hener το 1936[18] και αποτελεί κατηγοριοποίηση των συναισθημάτων σε 8 κλάσεις (βλέπε Σχήμα 1.1). Η κάθε κλάση σχετίζεται με την γειτονική της και διαφοροποιείται περισσότερο από την απέναντί της.
- Συνεχής Προσέγγιση (Dimensional Approach): Μια διαφορετική προσέγγιση στο πρόβλημα είναι η αναγνώριση των συναισθημάτων δεδομένου της θέσης τους ως προς άξονες που ορίζουν ένα χώρο διαστάσεων. Οι άξονες αυτοί αντιστοιχούν σε έννοιες που υποθετικά όταν συνδυαστούν ορίζουν το συνολικό εύρος των ανθρώπινων συναισθημάτων. Η σημαντικότερη συνεχής αναπαράσταση συναισθημάτων ορίζεται από το μοντέλο του Russell που αναπτύχθηκε το 1980[33]. Το μοντέλο του Russell αποτελείται από ένα διδιάστατο χώρο που ορίζεται από τους άξονες σθένους ή τερπνότητας (valence) και έντασης διέγερσης (arousal) (βλέπε Σχήμα 1.2). Ένα ακόμα δημοφιλές μοντέλο περιλαμβάνει επιπλέον και τον άξονα της επικράτησης (dominance) δημιουργώντας ένα τριδιάστατο επίπεδο.

Figure 1.1: Μοντέλο συναισθημάτων του Henver (από [18])



Figure 1.2: Μοντέλο συναισθημάτων του Russell (από [33])



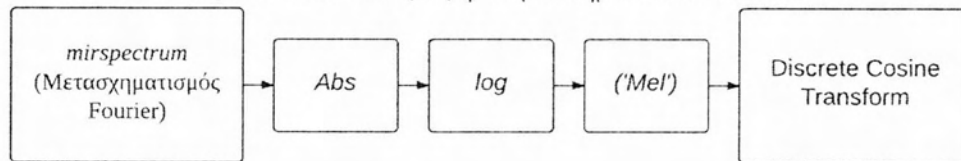
2 Εξαγωγή Χαρακτηριστικών

2.1 Εισαγωγή

Η εμπειρία της μουσικής ακρόασης μπορεί να θεωρηθεί μια πολυδιάστατη εμπειρία συναισθημάτων. Ο ακροατής μπορεί συνήθως να παρατηρήσει ότι υπάρχει ένα πρότυπο συσχέτισης μεταξύ πρόκλησης συναισθημάτων και διαφόρων ακουστικών χαρακτηριστικών. Για παράδειγμα, η συναισθηματική διέγερση (arousal) συνήθως σχετίζεται με το ρυθμό (αργός / γρήγορος), την τονικότητα (χαμηλή / υψηλή) και την ένταση (χαμηλή / υψηλή), ενώ το θετικό και αρνητικό σθένος του συναισθήματος (valence) σχετίζεται περισσότερο με την αρμονία (συμφωνία / παραφωνία) και το κλειδί (ματζόρε / μινόρε). Επίσης η συναισθηματική αντίληψη σπανίως συνδέεται με μόνο ένα μουσικό παράγοντα αλλά με ποικίλο συνδυασμό τους. Για παράδειγμα, ένα τραγούδι με γρήγορο ρυθμό και δυνατές μη αρμονικές συγχορδίες κατά πάσα πιθανότητα προκαλεί ένα συναίσθημα θυμού/έντασης, ενώ ένα τραγούδι με χαμηλόφωνες αρμονικές νότες αργού ρυθμού προκαλεί αίσθημα χαλάρωσης/ηρεμίας.

Όπως συνοψίζεται και στον Πίνακα 2.1 έχουν επιλεγεί διάφορα χαρακτηριστικά από 5 κατηγορίες για εξαγωγή τους από τα αρχεία ήχου της βάσης δεδομένων. Το κάθε ένα από αυτά αποτελεί ένα στοιχείο του μονοδιάστατου διανύσματος, το οποίο αντιπροσωπεύει κάθε δείγμα μουσικής που χρησιμοποιήθηκε. Τα χαρακτηριστικά που επιλέχθηκαν καλύπτουν όλο το φάσμα του τετραδιάστατου μοντέλου αντίληψης της μουσικής. Συγκεκριμένα, αποτελούνται από δυναμικά, ρυθμικά, αρμονικά και φασματικά. Κύρια έμφαση δόθηκε στα MFCC (Mel Frequency Cepstral Coefficients), τα οποία αποτελούν πάνω από το 50% του συνολικού διανύσματος χαρακτηριστικών και για αυτό το λόγο τα θεωρήσαμε ξεχωριστή κατηγορία. Η εξαγωγή των χαρακτηριστικών έγινε ανά παράθυρο (frame based) με μεταβλητό μέγεθος (ανάλογα με το είδος του χαρακτηριστικού). Από το σύνολο των frames χρησιμοποιήθηκαν ο μέσος όρος και η τυπική απόκλιση για μείωση του όγκου δεδομένων. Το συνολικό μέγεθος του διανύσματος χαρακτηριστικών κατέληξε σε 89 στοιχεία. Το διάνυσμα επιλέχθηκε με βάση δημοσιεύσεις διαφόρων ερευνητών Αναγνώρισης Συναισθήματος Μουσικής (Music Emotion Recognition - MER) με σημαντικότερο άξονα τη δημοσίευση των Saari και Eurola [34]. Για την εξαγωγή των χαρακτηριστικών χρησιμοποιήθηκε το MIRtoolbox (v 1.5)[27] για Matlab. Το παρόν κεφάλαιο περιγράφει τις σημασιολογικές έννοιες αυτών των χαρακτηριστικών αλλά και πως έγινε η εξαγωγή τους.

Figure 2.1: Εξαγωγή Χαρακτηριστικών MFCC



2.2 Mel Frequency Cepstral Coefficients (MFCC)

Τα MFCC είναι χαρακτηριστικά που παρέχουν φασματική πληροφορία σημάτων ήχου. Χρησιμοποιούνται ευρέως στην αναγνώριση φωνής, ομιλητή, αλλά και σε εφαρμογές Εξαγωγής Μουσικής Πληροφορίας (Music Information Retrieval - MIR) όπως αναγνώριση είδους μουσικής. Αυτό που καθιστά τα MFCC τόσο αποτελεσματικά είναι ότι οι ζώνες συχνότητων τοποθετούνται λογαριθμικά πάνω στην κλίμακα Mel (Mel scale), η οποία προσεγγίζει την απόκριση της ανθρώπινης ακοής πολύ καλύτερα από την γραμμική κλίμακα συχνότητων.

Σαν πρώτο βήμα για την εξαγωγή των συντελεστών MFCC υπολογίζεται ο μετασχηματισμός Fourier του σήματος. Στη συνέχεια, η απόλυτη τιμή του φάσματος που προκύπτει φιλτράρεται με ένα τριγωνικό filterbank που αντιστοιχεί στην κλίμακα Mel και το αποτέλεσμα μετασχηματίζεται με DCT (Discrete Cosine Transform), με αυτόν τον τρόπο συμπυκνώνεται το μεγαλύτερο ποσοστό της ενέργειας σε λίγους συντελεστές.

Για την εξαγωγή των MFCC χρησιμοποιήθηκε η συνάρτηση `mirmfcc()` σε frame decomposed σήμα με τετραγωνικό παράθυρο 50ms και επικάλυψη 50%. Επιλέχθηκαν οι 13 πρώτοι συντελεστές ανά παράθυρο και υπολογίστηκε ο μέσος όρος και η διασπορά αυτών. Επίσης, υπολογίστηκε ο μέσος όρος και η διασπορά του delta (1η παράγωγος) των MFCC. Στο Σχήμα 2.1 συνοψίζεται η διαδικασία (χαρακτηριστικά 1-52).

2.3 Δυναμικά Χαρακτηριστικά

Ένα από τα θεμελιώδη χαρακτηριστικά της μουσικής είναι η ένταση (loudness, intensity). Η δυνατής έντασης μουσική μπορεί να προκαλέσει ποικίλες εκφράσεις της συναισθηματικής έντασης όπως ενθουσιασμό, υπερένταση, θυμό και χαρά, από την άλλη η απαλή μουσική σχετίζεται με συναισθήματα ηρεμίας, τρυφερότητας, λύπης και φόβου. Τα χαρακτηριστικά που σχετίζονται με την ένταση τα κατηγοριοποιούμε στο σύνολο των δυναμικών χαρακτηριστικών. Τα χαρακτηριστικά αυτά σχετίζονται περισσότερο με το πλάτος του σήματος καθώς αυτό δηλώνει την αντίληψη έντασης ενός ακουστικού σήματος (perceived loudness). Η ένταση και η ακουστική ενέργεια ενός σήματος σχετίζεται άμεσα με το πλάτος του (amplitude). Το πλάτος ενός ακουστικού σήματος είναι η μέτρηση της διακύμανσης του μέσου (στο οποίο διαδίδεται) από την κατάσταση

ισορροπίας του. Για παράδειγμα, αυξάνοντας την ένταση σε ένα στερεοφωνικό σύστημα μπορούμε να παρατηρήσουμε την αύξηση της κίνησης του αέρα προς και από το ηχείο. Αυτό συμβαίνει γιατί αυξάνοντας το πλάτος, αυξήθηκε η διατάραξη των μορίων αέρα μπροστά από το ηχείο. Τα χαρακτηριστικά που επιλέχτηκαν για να καθοριστεί η δυναμικότητα ενός κομματιού είναι τα RMS Energy, Attack time, Attack slope και Low-energy rate.

2.3.1 Ενέργεια Ρίζας Μέσου Τετραγώνου

Υπολογίζοντας τον μέσο όρο του πλάτους ενός σήματος δεν μας δίνει απαραίτητα αρκετή πληροφορία για το πλάτος του. Αν, για παράδειγμα, έχουμε ένα ημιτονοειδές σήμα, ο μέσος όρος του πλάτους θα ισούται με μηδέν λόγω της συμμετρίας του σήματος γύρω από τον άξονα του χρόνου. Για αυτό το λόγο η μέτρηση του μέσου όρου του πλάτους συνήθως υπολογίζεται χρησιμοποιώντας τη Ρίζα Μέσου Τετραγώνου (Root Mean Square - RMS). Το RMS του πλάτους μας δίνει τη συνολική ενέργεια του σήματος και υπολογίζεται, παίρνοντας τη ρίζα του μέσου όρου του τετραγώνου του πλάτους.

$$x_{\text{rms}} = \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2} = \sqrt{\frac{1}{n} (x_1^2 + x_2^2 + \dots + x_n^2)},$$

οπου x_i το πλάτος κάθε δείγματος και n ο αριθμός των δειγμάτων.

Για τον υπολογισμό του RMS energy χρησιμοποιήθηκε η συνάρτηση `mirrms()` με `frame decomposed` σήμα τετραγωνικού παραθύρου 50ms και 50% επικάλυψη. Στη συνέχεια, υπολογίστηκε ο μέσος όρος και η τυπική απόκλιση των παραθύρων ανά απόσπασμα (χαρακτηριστικά 53-54).

2.3.2 Χρόνος Επίθεσης και Κλίση Επίθεσης

Στα μουσικά κομμάτια είναι σύνηθες να υπάρχουν σημεία υψηλότερης έντασης (πλάτους) τα οποία καθορίζουν συνήθως τα μουσικά γεγονότα του τραγουδιού (όπως οι νότες). Ο χρόνος επίθεσης (attack time) είναι ένα υψηλότερου επιπέδου χαρακτηριστικό το οποίο δηλώνει τη διάρκεια του κάθε ενός από τα attacks (δηλαδή των σημείων γρήγορης εναλλαγής πλάτους) (βλέπε Σχήμα 2.2).

Η κλίση επίθεσης (attack slope) μας δίνει την κλίση από το χαμηλότερο στο υψηλότερο πλάτος κατά τη διάρκεια των attacks (βλέπε Σχήμα 2.3).

Αυτά τα δύο χαρακτηριστικά μας περιγράφουν την ομαλότητα/αιχμηρότητα του φακέλου πλάτους (amplitude envelope). Ένας αιχμηρός φάκελος συνήθως προκαλεί συναισθήματα έκπληξης, θυμού, χορευτικής διάθεσης κ.ά., ενώ ένας ομαλός φάκελος, συναισθήματα που σχετίζονται με την ηρεμία.

Figure 2.2: Υπολογισμός χρόνου επίθεσης (από [26])

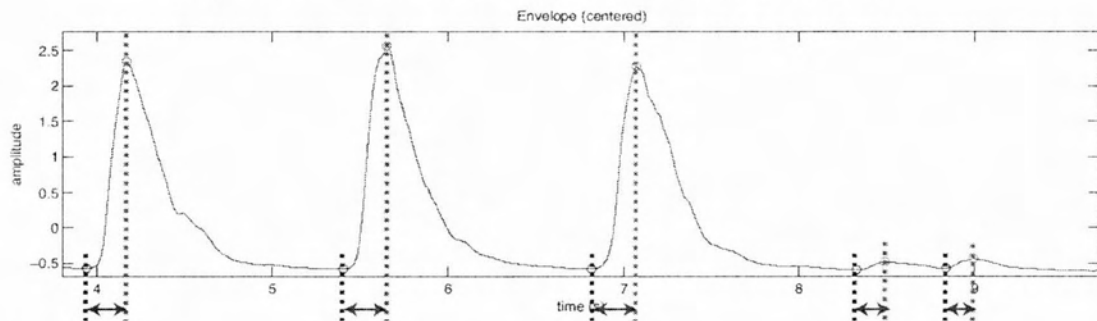
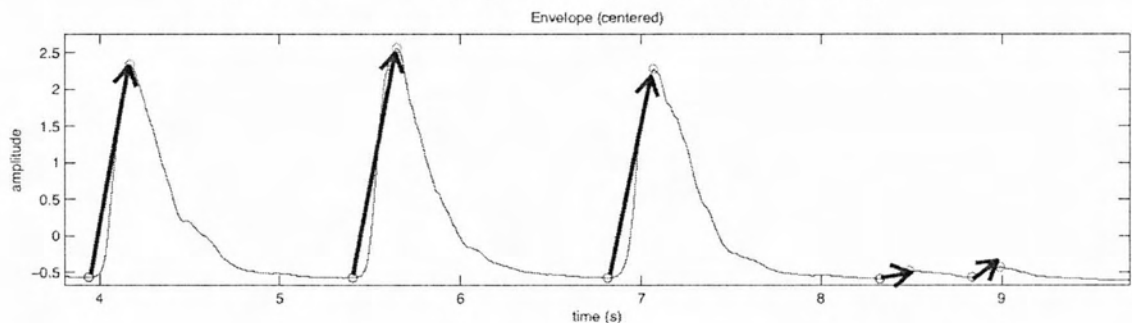


Figure 2.3: Υπολογισμός κλίσης επίθεσης (από [26])

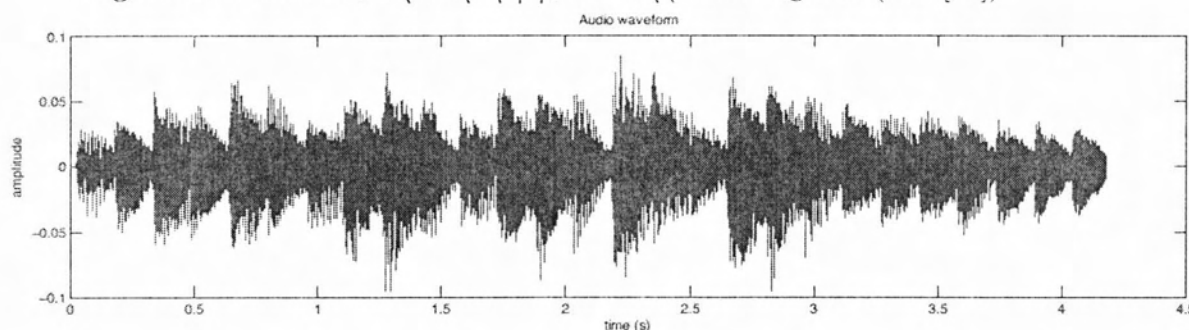


Για να υπολογιστεί το attack time σαν πρώτο βήμα πρέπει να υπολογιστεί ο φάκελος πλάτους της κυματομορφής. Πρακτικά ο φάκελος είναι το εξωτερικό περίγραμμα του σήματος και έχει μεγάλη χρησιμότητα στην ανίχνευση μουσικών γεγονότων όπως οι νότες. Παρακάτω περιγράφεται η διαδικασία εξαγωγής του φακέλου με χρήση φίλτρου. Πρώτα, οι αρνητικοί λοβοί του σήματος αντανακλώνται στο πεδίο των θετικών με χρήση της συνάρτησης απόλυτης τιμής (abs) του Matlab (full wave retriification). Στη συνέχεια, το σήμα περνά από ένα χαμηλοπερατό φίλτρο το οποίο εξομαλύνει τα σημεία με μεγάλες διακυμάνσεις και κρατάει από το σήμα τη μακροπρόθεσμη εξέλιξη του. Το (default) φίλτρο που χρησιμοποιείται είναι ένα IIR με συντελεστή οπισθοδρόμησης. Τέλος, το εξομαλυνμένο σήμα δειγματοληπτείται με ρυθμό υποδειγματοληψίας N (default $N=16$), καθώς υπάρχει πλέον υψηλός συσχετισμός μεταξύ συνεχόμενων δειγμάτων (βλέπε Σχήμα 2.4).

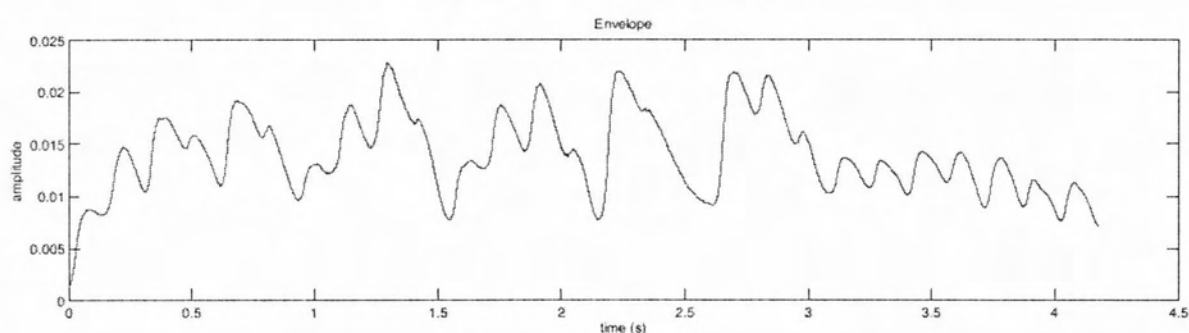
Το επόμενο βήμα είναι η ανίχνευση των μουσικών γεγονότων. Σε αυτή την φάση επιλέγονται τα διάφορα τοπικά μέγιστα (peaks) πάνω στην κυματομορφή του φακέλου. Κάθε ένα από αυτά αντιστοιχεί σε ένα γεγονός. Τέλος, μετράται ο χρόνος από τη στιγμή που το πλάτος αρχίζει να ανεβαίνει μέχρι και το peak. Αυτό το χαρακτηριστικό είναι ο χρόνος επίθεσης (attack time) ο οποίος μας δείχνει την ομαλότητα με την οποία παίζονται οι νότες.

Άλλο ένα χαρακτηριστικό της φάσης επίθεσης που χρησιμοποιήθηκε είναι η κλίση

Figure 2.4: Φάκελος κυματομορφής του κομματιού ragtime (από [26])



Audio waveform of ragtime excerpt



Corresponding envelope of the ragtime excerpt

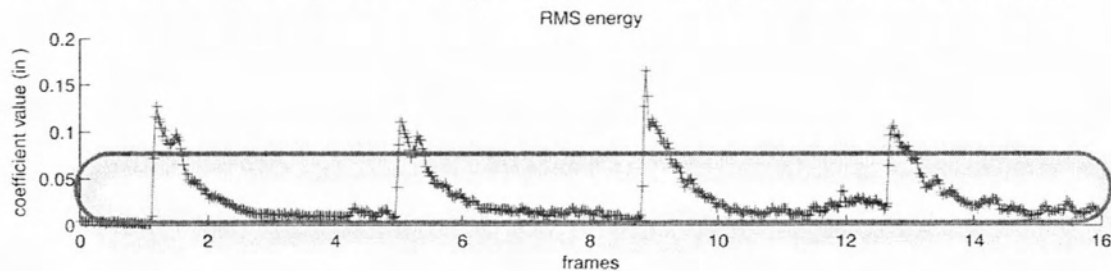
επίθεσης (attack slope), που πρακτικά μας δίνει την κλίση του ευθύγραμμου τμήματος το οποίο σχηματίζεται μεταξύ του αρχικού σημείου της φάσης επίθεσης και του τοπικού μέγιστου.

Ο μέσος όρος και η τυπική απόκλιση αυτών των τιμών χρησιμοποιήθηκαν στο διάγραμμα χαρακτηριστικών (χαρακτηριστικά 55 με 58).

2.3.3 Ποσοστό Χαμηλής Ενέργειας

Το RMS του πλάτους δεν μας δίνει απαραίτητα αρκετή πληροφορία για την κατανομή της ενέργειας σε ένα δείγμα. Χρησιμοποιώντας την κυματομορφή της ενέργειας μπορούμε να αποφανθούμε εάν η ενέργεια είναι συνεπής καθ' όλη τη διάρκεια του σήματος ή εάν εμφανίζονται ασυνέπειες. Ένας τρόπος να κάνουμε αυτή την εκτίμηση είναι να υπολογίσουμε το ποσοστό χαμηλής ενέργειας (low energy rate), δηλαδή το ποσοστό των σημείων που εμφανίζουν μικρότερη του μέσου όρου ενέργεια.

Για παράδειγμα, στο Σχήμα 2.5 μπορούμε να παρατηρήσουμε ότι λόγω κάποιων σπάνιων frames που εμφανίζουν υψηλή ενέργεια σε σχέση με το υπόλοιπο σήμα, το μεγαλύτερο ποσοστό των frames είναι κάτω από το κατώφλι του μέσου όρου και πράγματι, εάν

Figure 2.5: Παράδειγμα υπολογισμού Ποσοστού Χαμηλής Ενέργειας (από [26])

υπολογίσουμε το low-energy rate παίρνουμε την τιμή 0.71317.

Ο μέσος όρος ανά τετραγωνικό παράθυρο 2s με 50% επικάλυψη του Low Energy Ratio χρησιμοποιήθηκε στο διάνυσμα χαρακτηριστικών (χαρακτηριστικό 59).

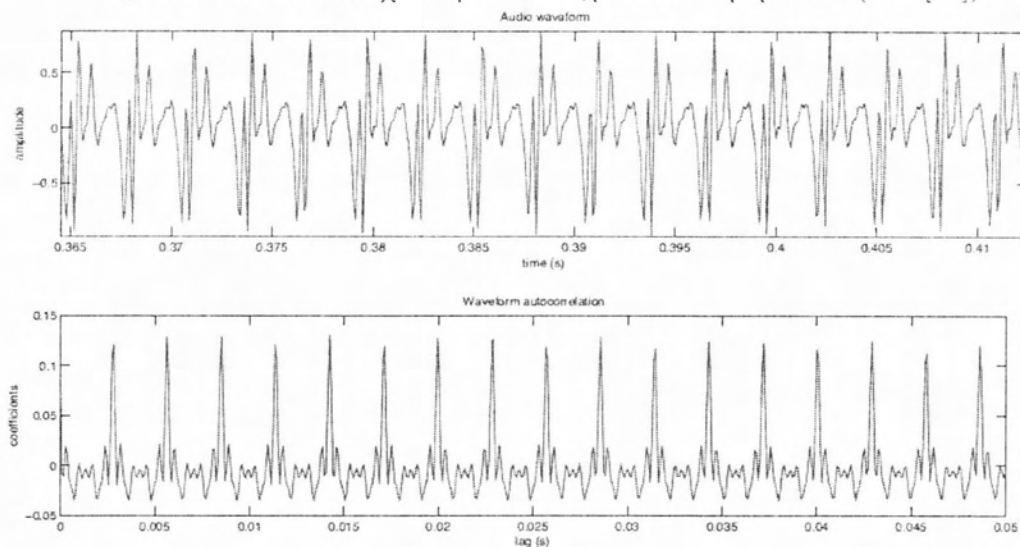
2.4 Ρυθμικά Χαρακτηριστικά

Ως ρυθμός μπορεί γενικά να οριστεί “οποιαδήποτε κίνηση που χαρακτηρίζεται από την οργανωμένη διαδοχή ισχυρών και αδύναμων στοιχείων”. Στη μουσική ο ρυθμός εκφράζει την οργανωμένη διαδοχή των γεγονότων που χρησιμοποιούνται, δηλαδή των ήχων και των σιωπών. Ένας σταθερός (κανονικός) ρυθμός μπορεί να γίνει αντιληπτός σαν έκφραση χαράς και ηρεμίας, ενώ ένας μεταβαλλόμενος (τραχύς) ρυθμός προκαλεί συναισθήματα έκπληξης, ανησυχίας, φόβου κ.λ.π. Ο ρυθμός ενός μουσικού κομματιού καθορίζεται από την ταχύτητα (tempo), δηλαδή από το πόσο γρήγορα ή αργά διαδέχεται ένας παλμός (ή χτύπος) τον επόμενο. Η ταχύτητα στη μουσική μετράται συχνά σε παλμούς ανά λεπτό (beats per minute - bpm), δηλαδή κάνοντας λόγο για ‘60 bpm’, εννοούμε ‘ταχύτητα που ισοδυναμεί με εξήντα παλμούς το λεπτό’. Το tempo θεωρείται ο σημαντικότερος παράγοντας συναισθηματικής έκφρασης της μουσικής. Ένα γρήγορο κομμάτι μπορεί να προκαλέσει αυξημένη διάθεση για δραστηριότητα, έξαψη, χαρά, ανησυχία, θυμό, φόβο, ενώ ένα κομμάτι με αργό tempo συνήθως δημιουργεί συναισθήματα ηρεμίας, λύπης, τρυφερότητας ή ανίας. Τα ρυθμικά χαρακτηριστικά που επιλέχθηκαν είναι τα tempo change, metrical centroid, metrical strength, pulse clarity και event density, τα οποία επιλέχθηκαν με βάση την πρόταση του [24].

2.4.1 Καμπύλη Ανίχνευσης Γεγονότων

Σαν πρώτο βήμα στην ανάλυση ρυθμού ενός κομματιού πρέπει να υπολογιστεί η καμπύλη ανίχνευσης γεγονότων (onset detection curve). Τα μουσικά γεγονότα σε αυτήν την καμπύλη υποδεικνύονται από τα τοπικά μέγιστα (peaks), όπου το ύψος του κάθε ενός σχετίζεται με την ενεργειακή (ή/και φασματική) σημασία του γεγονότος. Για την ανίχνευση της καμπύλης γεγονότων μπορούμε να ακολουθήσουμε δύο βασικές κατευθύνσεις. Η πρώτη είναι να υπολογίσουμε αρχικά τον φάκελο πλάτους (βλέπε

Figure 2.6: Αυτοσυσχέτιση ενός στιγμιότυπου τρομπέτας (από [26])



Ενότητα 2.3.2) που μας δίνει την γενική εξέλιξη της ενέργειας στο χρόνο, ενώ η δεύτερη να υπολογίσουμε τη φασματική ροή (spectral flux), δηλαδή να αξιολογήσουμε την απόσταση μεταξύ διαδοχικών στιγμών σε σχέση με τη φασματική κατανομή.

2.4.2 Αυτοσυσχέτιση

Ο ρυθμός σχετίζεται με την περιοδικότητα ενός σήματος. Για το λόγο αυτό πολύ χρήσιμη για τη μελέτη του ρυθμού είναι η συνάρτηση αυτοσυσχέτισης (autocorrelation). Η αυτοσυσχέτιση ενός σήματος υπολογίζεται από τη σχέση:

$$R_{xx}(j) = \sum_n x_n \bar{x}_{n-j},$$

όπου x_n το πλάτος του σήματος τη χρονική στιγμή n , ενώ \bar{x}_{n-j} το πλάτος του σήματος τη χρονική στιγμή $n-j$.

Για μία δεδομένη καθυστέρηση (lag) j , η αυτοσυσχέτιση ισούται με το γινόμενο σημείο προς σημείο του σήματος με μία μετακινημένη εκδοχή του σήματος κατά j . Για κάθε καθυστέρηση j παίρνουμε την καμπύλη αυτοσυσχέτισης όπου οι κορυφές (peaks), μας δίνουν τις καθυστερήσεις στις οποίες το σήμα εμφανίζει μεγαλύτερη περιοδικότητα (βλέπε Σχήμα 2.6).

2.4.3 Υπολογισμός Περιοδικότητας

Ο ρυθμός ενός κομματιού σχετίζεται με την περιοδικότητα στην αλληλουχία των peaks της καμπύλη γεγονότων. Αυτή η περιοδικότητα μπορεί να ανιχνευτεί μέσα από τον

υπολογισμό της συνάρτησης αυτοσυσχέτισης σε συνεχόμενα μεγάλα frames (μερικών δευτερολέπτων) στην καμπύλη γεγονότων. Για παράδειγμα, εάν έχουμε ένα κομμάτι με tempo 120 bpm, δηλαδή 2 παλμών ανά δευτερόλεπτο, η συνάρτηση αυτοσυσχέτισης θα μας δώσει υψηλή αυτοσυσχέτιση μετά από 0.5 s, όπως επίσης και μετά από 1 s, 1.5 s κλπ. Ένας τρόπος λοιπόν να εξάγουμε το tempo είναι να επιλέξουμε το υψηλότερο peak από ένα φάσμα επαναλαμβανόμενων παλμών (συνήθως μεταξύ 40 και 200 bpm) και να το θεωρήσουμε ως το επικρατέστερο. Ο κύριος περιορισμός αυτής της μεθόδου είναι ότι, εάν στην τοπική εξέλιξη ενός κομματιού δίνεται έμφαση σε διαφορετικά μουσικά μέτρα, η ανίχνευση tempo θα μεταβαίνει συνέχεια από ένα bpm σε άλλο (στην μισή ταχύτητα ή στην διπλάσια ταχύτητα). Αυτό το πρόβλημα είναι πολύ πιθανό καθώς στη μουσική οι μεταβάσεις σε μισής (ή διπλάσιας) ταχύτητας γεγονότα είναι συνηθισμένες. Μία λύση σε αυτό το πρόβλημα είναι η ανίχνευση πολλαπλών μέτρων και η αξιολόγηση τους σύμφωνα με την ισχύ τους στην διάρκεια του δείγματος. Τα διάφορα μέτρα τοποθετούνται σύμφωνα με την ισχύ τους σε επίπεδα τα οποία ονομάζονται μετρικά επίπεδα (metrical levels).

2.4.4 Τέμπο και Αλλαγή Τέμπο

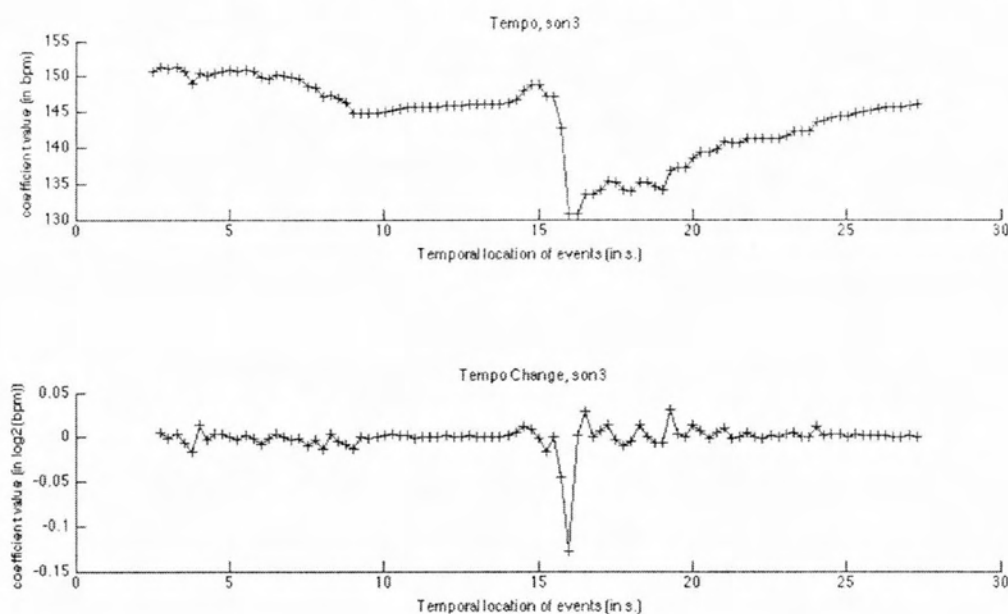
Όταν ολοκληρωθεί η κατασκευή της μετρικής δομής, ένα επίπεδο επιλέγεται ως το κυρίαρχο μετρικό επίπεδο. Αυτή η αξιολόγηση γίνεται σύμφωνα με το μεγαλύτερο συνολικό άθροισμα των σκορ της αυτοσυσχέτισης ανά μετρικό επίπεδο ξεχωριστά. Πάνω στο κυρίαρχο μετρικό επίπεδο επιλέγεται το tempo, δηλαδή ο αριθμός των peaks (στο συγκεκριμένο μετρικό επίπεδο) στην διάρκεια ενός λεπτού (bpm). Βέβαια η επιλογή του κυρίαρχου μέτρου παραμένει υποκειμενική και το tempo αυτό καθ' αυτό δεν μας δίνει απαραίτητα αρκετή πληροφορία για την εμπειρία της μουσικής ακρόασης. Αντί αυτού ως χαρακτηριστικό στην παρούσα εργασία επιλέχθηκε η αλλαγή του tempo στην καθορισμένη διάρκεια (tempo change). Η αλλαγή του tempo υπολογίζει την αλλαγή στο tempo ανάμεσα σε διαδοχικά frames εκφρασμένη σε λογαριθμική κλίμακα (βλέπε Σχήμα 2.7).

Τα χαρακτηριστικά 60 - 61 είναι ο μέσος όρος και η τυπική απόκλιση του tempo change για frame decomposed σήματα τετραγωνικού παραθύρου 2s με hop 50%.

2.4.5 Μετρικό κεντροειδές και Μετρική Ισχύς

Ένα σημαντικό στοιχείο αντίληψης του ρυθμού είναι ότι ορισμένα μετρικά επίπεδα είναι πιο ισχυρά από άλλα. Η μέθοδος που επιλέχθηκε για την εξαγωγή αυτού του χαρακτηριστικού είναι ο υπολογισμός του κεντροειδούς (centroid) ενός πλήθους μετρικών επιπέδων. Τα επίπεδα αυτά επιλέγονται σύμφωνα με την ισχύ τους, δηλαδή το σκορ της αυτοσυσχέτισης. Για κάθε frame επιλέγεται το κυρίαρχο μετρικό επίπεδο και υπολογίζεται το κεντροειδές χρησιμοποιώντας ως βάρος το αντίστοιχο σκορ της αυτοσυσχέτισης. Η συνολική καμπύλη των κεντροειδών αναπαριστά την εξέλιξη της μετρικής δραστηριότητας στη διάρκεια του κομματιού.

Figure 2.7: Τέμπο και Αλλαγή Τέμπο (από [26])



Ένα ακόμα σημαντικό χαρακτηριστικό της μετρικής δραστηριότητας είναι η ισχύς του κυρίαρχου μέτρου. Ένα κομμάτι με ισχυρό κυρίαρχο μέτρο δείχνει ότι έχει ξεκάθαρο ρυθμικό τονισμό, ενώ χαμηλή μετρική ισχύς δείχνει ένα κομμάτι με πιο αφηρημένο (τραχύ) ρυθμό. Η μετρική ισχύς υπολογίζεται απλά ως το μεγαλύτερο άθροισμα των σκωρ αυτοσυσχέτισης από όλα τα μετρικά επίπεδα (βλέπε Σχήμα 2.8).

Τα χαρακτηριστικά 62 - 65 είναι ο μέσος όρος και η τυπική απόκλιση του metrical centroid και metrical strength για frame decomposed σήματα τετραγωνικού παραθύρου 2s με hop 50%.

2.4.6 Καθαρότητα Παλμών

Η καθαρότητα των παλμών (pulse clarity) είναι ένα υψηλού επιπέδου χαρακτηριστικό το οποίο δηλώνει πόσο εύκολα γίνεται αντιληπτός ο ρυθμός ενός κομματιού. Υπάρχουν διάφορες κατευθύνσεις για τον υπολογισμό του pulse clarity. Στην παρούσα εργασία χρησιμοποιήθηκε η πρόταση των Olivier Lartillot, Tuomas Eerola για εξαγωγή του pulse clarity μέσω της ανάλυσης συνάρτησης αυτοσυσχέτισης του κομματιού [25]. Συγκεκριμένα, ως καθαρότητα του παλμού παίρνουμε το μέγιστο της καμπύλης αυτοσυσχέτισης του φακέλου πλάτους του σήματος (βλέπε Σχήμα 2.9).

Το χαρακτηριστικό 66 είναι το pulse clarity για όλη τη διάρκεια του σήματος.

Figure 2.8: Μετρικό κεντροειδές και Μετρική ισχύς (από [26])

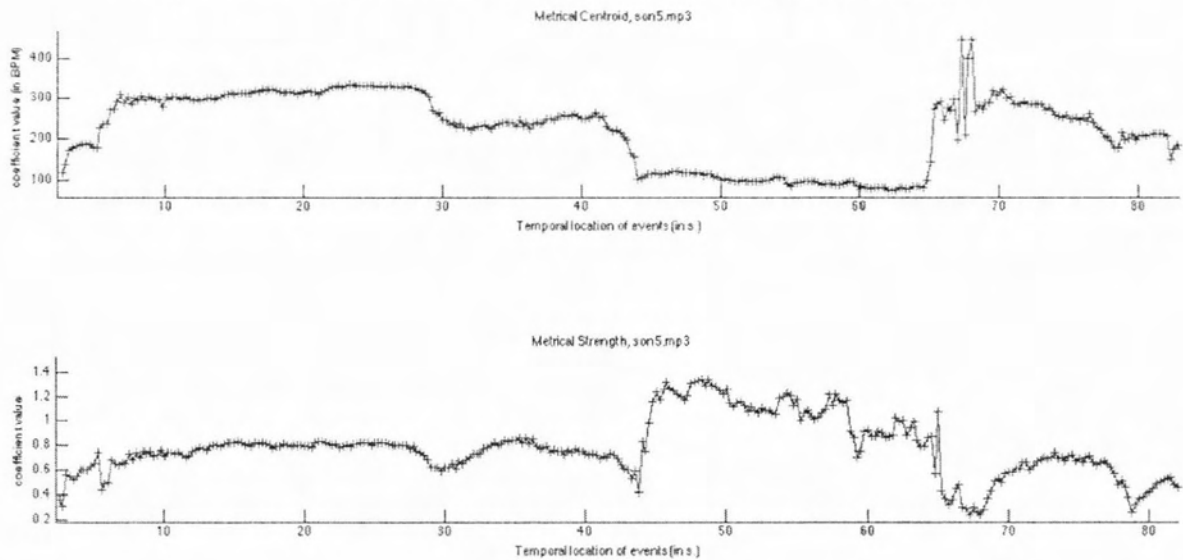
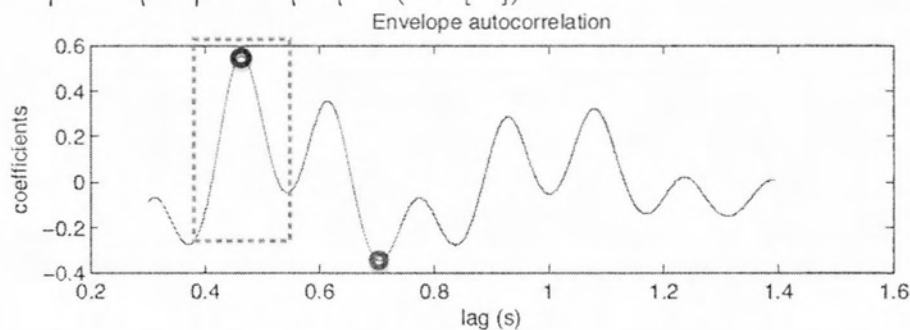


Figure 2.9: Το μέγιστο (μαύρος κύκλος) της συνάρτησης αυτοσυσχέτισης μας δίνει την καθαρότητα του ρυθμού (από [25])



2.4.7 Πυκνότητα Γεγονότων

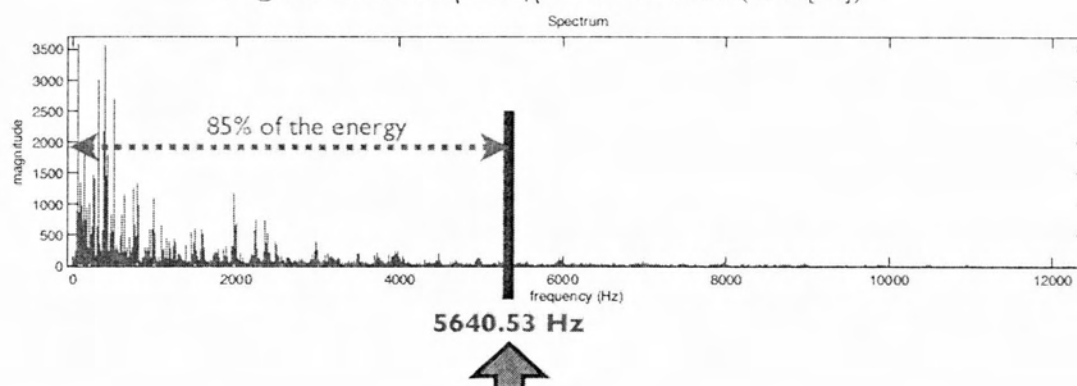
Τέλος, ως ένδειξη της ταχύτητας ενός κομματιού μπορούμε να χρησιμοποιήσουμε τη μέση συχνότητα των γεγονότων ανά frame από την καμπύλη γεγονότων. Αυτό το χαρακτηριστικό ονομάζεται πυκνότητα γεγονότων (event density).

Το χαρακτηριστικό 67 είναι το event density για όλη τη διάρκεια του σήματος.

2.5 Φασματικά Χαρακτηριστικά

Τα σημαντικότερα φασματικά χαρακτηριστικά όπως ήδη αναφέρθηκε είναι τα MFCC, όμως διαφορετικές αναλύσεις πάνω στο φάσμα της κυματομορφής μπορούν να μας

Figure 2.10: Παράδειγμα roll-off 85% (από [26])



δώσουν χρήσιμη πληροφορία για τη χροιά και το χρώμα ενός κομματιού. Παρακάτω δίνεται περιγραφή και ανάλυση αυτών των χαρακτηριστικών.

Για την εξαγωγή των παρακάτω χαρακτηριστικών τα αποσπάσματα χωρίστηκαν σε παράθυρα 46ms με 50% επικάλυψη. Σαν τελικά χαρακτηριστικά χρησιμοποιήθηκαν ο μέσος όρος και η τυπική απόκλιση των zero crossing rate, rolloff85, centroid, entropy και irregularity, ενώ μόνο ο μέσος όρος χρησιμοποιήθηκε στα spread, skewness, flux και flatness (χαρακτηριστικά 68 - 81).

2.5.1 Ρυθμός Μηδενικής Διασταύρωσης

Το zero crossing rate είναι ένα χαρακτηριστικό που μας δείχνει πόσες φορές η κυματομορφή πέρασε τον άξονα του χρόνου (άλλαξε πρόσημο). Το zero-cross rate μιας κυματομορφής χρησιμοποιείται συχνά για τη διάκριση μεταξύ θορύβου, ομιλίας και μουσικής. Συνήθως, η τιμή του είναι υψηλή για θόρυβο και ομιλία, μέτρια για μουσική με φωνητικά και χαμηλή για ορχηστρική μουσική.

2.5.2 Φασματική Κύλιση

Ένας από τους τρόπους για την εκτίμηση της ποσότητας των υψηλών συχνοτήτων ενός σήματος είναι να εντοπίσουμε την συχνότητα που ένα συγκεκριμένο ποσοστό της συνολικής ενέργειας περιέχεται από αυτή τη συχνότητα και κάτω. Αυτό το ποσοστό είναι συνήθως το 85% (σταθεροποιήθηκε από τους Tzanetakis και Cook, 2002). Αυτό το χαρακτηριστικό ονομάζεται roll-off και είναι μια εκ των δύο εναλλακτικών για τον εντοπισμό της ενέργειας υψηλών συχνοτήτων έναντι της μεθόδου που κρατάει μία σταθερή συχνότητα και εντοπίζει την ενέργεια κάτω από αυτή (βλέπε Σχήμα 2.10).

2.5.3 Φασματικό Κεντροειδές

Το φασματικό κεντροειδές (spectral centroid) δείχνει που βρίσκεται το κέντρο βάρους ενός φάσματος και έχει άμεση σχέση με τη φωτεινότητα (brightness) ενός ακουστικού σήματος. Έστω, $X_i(k)$, $k = 1 \dots, N$, οι συντελεστές του διακριτού μετασχηματισμού Fourier (DFT) του i -οστού βραχυπρόθεσμου παραθύρου, όπου N είναι το μήκος του παραθύρου. Το φασματικό κεντροειδές C_i του i -οστού παραθύρου ορίζεται ως το κέντρο βάρους του φάσματος, δηλαδή:

$$C_i = \frac{\sum_{k=1}^N (k+1)X_i(k)}{\sum_{k=1}^N X_i(k)}$$

2.5.4 Φασματική Έκταση

Η φασματική έκταση (spectral spread) μας δίνει το εύρος ζώνης του φάσματος. Πρακτικά περιγράφει τη διασπορά του φάσματος γύρω από το κεντροειδές. Υψηλή φασματική έκταση είναι αλληλένδετη με μεγάλο εύρος συχνοτήτων σε ένα κομμάτι.

2.5.5 Φασματική Λοξότητα

Η φασματική λοξότητα (spectral skewness) περιγράφει τη συμμετρία του φάσματος συχνοτήτων. Θετική τιμή της λοξότητας δείχνει ότι το κομμάτι τείνει προς τις υψηλές συχνότητες ενώ αρνητική προς τις χαμηλές.

2.5.6 Εντροπία

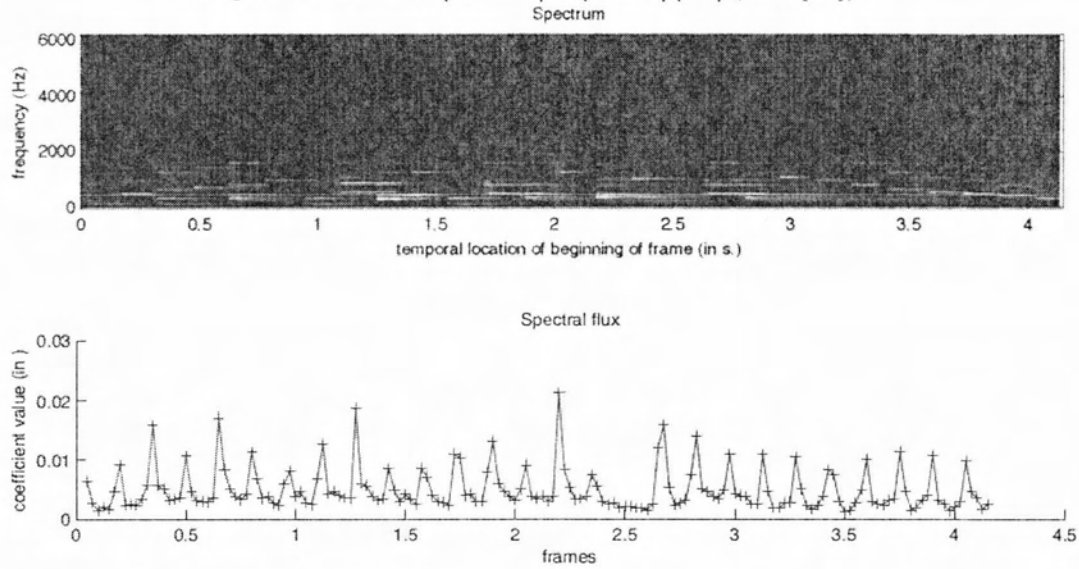
Στη θεωρία πληροφορίας η εντροπία είναι ένα μέγεθος το οποίο ορίζει την ποσότητα αβεβαιότητας μιας τυχαίας μεταβλητής. Υπολογίζεται από την εξίσωση του Shannon:

$$H(X) = - \sum_{i=1}^n p(x_i) \log_b p(x_i),$$

όπου $p(x_i)$ είναι η συνάρτηση πυκνότητας πιθανότητας της μεταβλητής x_i .

Ο υπολογισμός της εντροπίας σε ένα φάσμα συχνοτήτων μας δίνει την προβλεψιμότητα του φάσματος. Για παράδειγμα, εάν έχουμε ένα φάσμα που είναι εντελώς επίπεδο, η εντροπία του θα είναι μέγιστη καθώς μοντελοποιείται σε μία κατάσταση όπου έχουμε ίδια πιθανότητα για κάθε συχνότητα (άρα ελάχιστη προβλεψιμότητα). Αντίθετα, εάν έχουμε ένα φάσμα με μία κορυφή που κυριαρχεί έναντι των άλλων, τότε θα έχουμε ελάχιστη εντροπία, καθώς θα περιμένουμε το μεγαλύτερο μέρος των σημείων να κινείται σε αυτή τη συχνότητα.

Figure 2.11: Φάσμα και φασματική ροή (από [26])



2.5.7 Φασματική Ροή

Η φασματική ροή (spectral flux) είναι ένα μέγεθος το οποίο μετράει πόσο γρήγορα το φάσμα ενός σήματος αλλάζει. Μπορούμε να υπολογίσουμε τη φασματική ροή ως την απόσταση των συχνοτήτων μεταξύ διαδοχικών frames (βλέπε Σχήμα 2.11).

2.5.8 Φασματική ομαλότητα

Η φασματική ομαλότητα (spectral flatness / wiener entropy) περιγράφει εάν ένα φάσμα είναι ομαλό ή οδοντωτό (spiky). Ένα ομαλό φάσμα σημαίνει ότι περιέχει σε ίδιο βαθμό όλες τις συχνότητες οπότε προσεγγίζει το λευκό θόρυβο, ενώ ένα οδοντωτό δείχνει ότι η περισσότερη ενέργεια περιέχεται σε μικρο εύρος ζώνης συχνοτήτων και άρα προσεγγίζει περισσότερο μια μίξη ημιτονοειδών σημάτων. Υπολογίζεται από την αναλογία του γεωμετρικού και του αριθμητικού μέσου όρου:

$$\text{Flatness} = \frac{\sqrt[N]{\prod_{n=0}^{N-1} x(n)}}{\frac{\sum_{n=0}^{N-1} x(n)}{N}} = \frac{\exp\left(\frac{1}{N} \sum_{n=0}^{N-1} \ln x_n\right)}{\frac{1}{N} \sum_{n=0}^{N-1} x_n}$$

2.5.9 Παρατυπία

Η παρατυπία (irregularity) ενός φάσματος δηλώνει το βαθμό της μεταβολής του φάσματος μεταξύ διαδοχικών κορυφών. Υπολογίζεται ως το συνολικό άθροισμα του τετραγώνου της διαφοράς του πλάτους γειτονικών κορυφών (Jensen 1999):

$$R_m = \sum_{k=1}^N (a_k - a_{k+1})^2 / \sum_{k=1}^N a_k^2,$$

όπου a_k το πλάτος της k -ιοστής κορυφής και N το σύνολο των κορυφών.

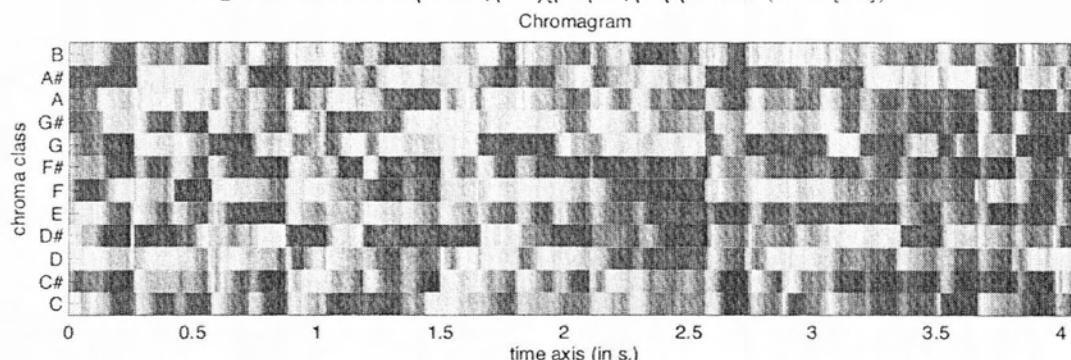
2.6 Αρμονικά Χαρακτηριστικά

Ένα από τα πιο σημαντικά χαρακτηριστικά στην εμπειρία της μουσικής ακρόασης είναι η αρμονία. Ως αρμονία ορίζουμε τον διαφορετικό τρόπο που συνδέονται οι διαφορετικές νότες (pitches) ή συγχορδίες. Οι συγχορδίες είναι σετ διαφορετικών νοτών (δύο ή περισσότερες) που παίζονται ταυτόχρονα. Η βάση κάθε συγχορδίας είναι μια κλίμακα, δηλαδή μια αλληλουχία τόνων που ανεβαίνουν σε συχνότητα. Στη θεωρία της δυτικής μουσικής υπάρχουν δύο θεμελιώδεις τύποι συγχορδιών, οι μείζονες συγχορδίες (ματζόρε) και οι ελάσσονες (μινόρε). Αυτές οι συγχορδίες αποτελούνται από 3 νότες (την 1η, την 3η και την 5η της κλίμακας). Η 1η ονομάζεται Τονική και δίνει το όνομα της συγχορδίας. Η 3η (Μέση) καθορίζει εάν η συγχορδία είναι μείζονα ή ελάσσονα. Η 5η ονομάζεται Δεσπόζουσα. Οι μείζονες τείνουν να δημιουργούν θετικά συναισθήματα (χαρά, ευθυμία, κ.ά.), ενώ οι ελάσσονες συνήθως τείνουν προς την αντίθετη κατεύθυνση (λύπη, ένταση, κ.ά.). Παρακάτω περιγράφονται διάφορα χαρακτηριστικά που μπορούμε να εξάγουμε σχετικά με την αρμονία ενός ήχου όπως η τονικότητα (key), δηλαδή τον τόνο που προσεγγίζει, ή όπως τον τρόπο (mode), δηλαδή αν προσεγγίζει τις μείζονες ή τις ελάσσονες συγχορδίες. Επίσης, περιγράφεται πως μπορούμε να εξάγουμε την τραχύτητα ενός ήχου και την δυσαρμονία του.

2.6.1 Χρωμόγραμμα

Σαν πρώτο βήμα για την προσέγγιση της τονικότητας ενός κομματιού μπορούμε να υπολογίσουμε το χρωμόγραμμα του. Το χρωμόγραμμα (Chromagram or Harmonic Pitch Class Profile - HPCP) δείχνει την κατανομή της ενέργειας σε συγκεκριμένους μουσικούς τόνους. Συγκεκριμένα είναι η προβολή του φάσματος πάνω σε 12 ξεχωριστά επίπεδα που αναπαριστούν τα 12 ημιτόνια της ισομερώς κατανεμημένης οκτάβας της δυτικής μουσικής. Στο χρωμόγραμμα δε λαμβάνεται υπόψη η απόλυτη συχνότητα αλλά η κατανομή της στα ημιτόνια της οκτάβας. Για παράδειγμα, frames με συχνότητα γύρω στα 440 Hz (C4) και 880 Hz (C5) προβάλλονται στο χρωμόγραμμα πάνω στον τόνο Ντο (C). Για τον υπολογισμό του χρωμογράμματος υπολογίζεται το φάσμα σε λογαριθμική κλίμακα με επιλογή των 20 υψηλότερων dB και με περιορισμό συγκεκριμένου εύρους συχνοτήτων που αντιστοιχούν σε ακέραιο αριθμό τόνων (12 ημιτόνια). Πριν από τον υπολογισμό του FFT προηγείται κανονικοποίηση της κυματομορφής (βλέπε Σχήμα 2.12).

Figure 2.12: Παράδειγμα χρωμογράμματος (από [26])



2.6.2 Ισχύς Τονικότητας, Καθαρότητα Τονικότητας και Τρόπος

Όπως αναφέρθηκε στον πρόλογο, στη δυτική μουσική υπάρχουν δύο βασικές κατηγοριοποιήσεις συγχορδιών, οι Μείζονες ή Ματζόρε (Major) και οι Ελάσσονες ή Μινόρε (Minor). Η ισχύς της τονικότητας (key strength) μας δίνει ένα σκορ μεταξύ -1 και 1 για κάθε μια από τα δύο είδη συγχορδιών, καθώς και τον επικρατέστερο τόνο (από τα 12 ημιτόνια). Η ισχύς της τονικότητας υπολογίζεται συγκρίνοντας το χρωμόγραμμα με 24 μινόρε και ματζόρε προφίλ (C Major, C Minor, C# Major, C# minor, D Major, κ.ο.κ.). Συγκεκριμένα [16], υπολογίζεται η διασυσχέτιση (cross-correlation) του κανονικοποιημένου χρωμογράμματος με αντίστοιχα προφίλ τα οποία περιέχουν 12 τιμές που αντιστοιχούν σε βαθμολόγηση κάθε τόνου ανά συγχορδία (βλέπε Σχήμα 2.13).

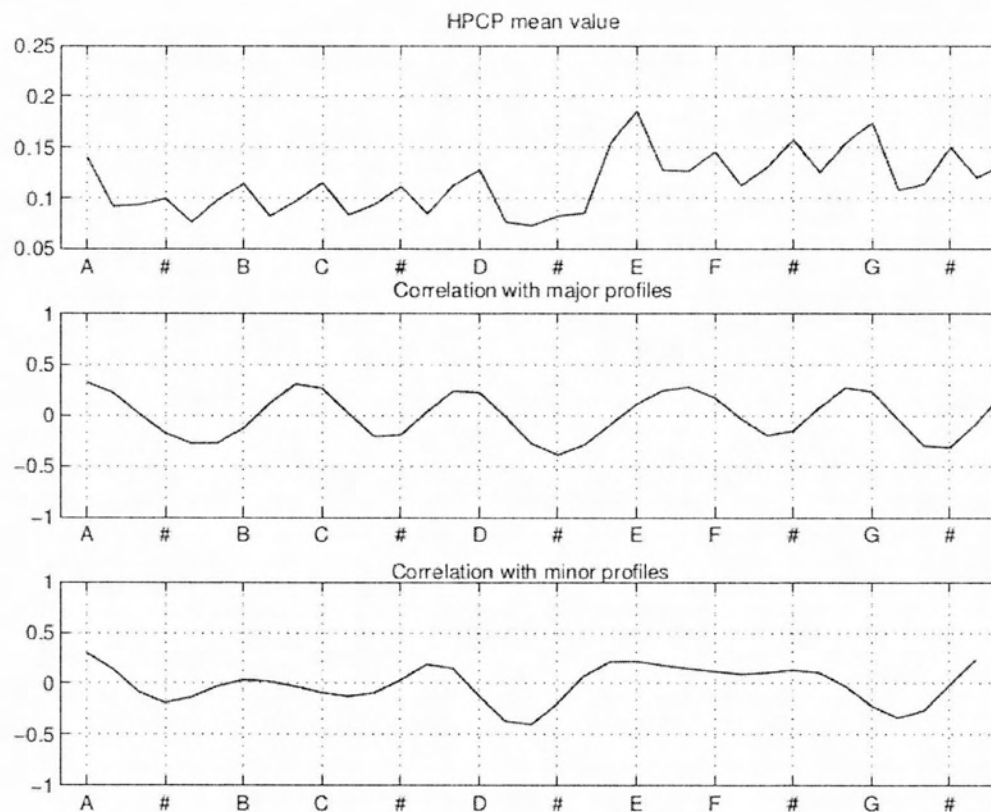
$$\text{KeyStrength}(i, j) = r(\text{HPCP}, K(i, j)),$$

όπου HPCP είναι το χρωμόγραμμα, $K(i, j)$ είναι το προφίλ της κλίμακας, $i = 1, 2$, (όπου 1 αναπαριστά το Ματζόρε και 2 το Μινόρε προφίλ), $j = 1 \dots 12$ για τις 12 πιθανές τονικές.

Υπολογίζοντας το μέγιστο και των δύο καμπυλών του key strength ανά στιγμή μπορούμε να υποθέσουμε την κυρίαρχη τονικότητα στη δεδομένη στιγμή (dominant key). Ακολουθώντας την ισχύ της κυρίαρχης τονικότητας στη διάρκεια του frame παίρνουμε την καμπύλη της διαύγειας της συγχορδίας (key clarity). Το key clarity μας δίνει πόσο “καθαρά” διακρίνονται οι συγχορδίες σε ένα frame (βλέπε Σχήμα 2.14).

Ο τρόπος (mode ή αλλιώς majorness) ενός αποσπάσματος αναπαριστά μια βαθμολόγηση μεταξύ -1 και 1 για το πόσο ματζόρε είναι ένα κομμάτι. Όσο πιο κοντά στο 1 τόσο πιο ματζόρε είναι, ενώ όσο πλησιάζει στο -1 τόσο πιο μινόρε. Από default η στρατηγική που ακολουθείται είναι να υπολογιστεί η υψηλότερη τιμή στην καμπύλη του ματζόρε key strength και αντίστοιχα του μινόρε key strength και στην συνέχεια να υπολογιστεί η διαφορά τους.

Figure 2.13: Παράδειγμα διασυσχέτισης ισχύος κλειδιού: String Quartet Op. 30 I Moderato, from Arnold Schoenberg (από [16])



2.6.3 Τονικό Κεντροειδές και Hcdf

Στη δυτική μουσική μία συγχορδία ορίζεται (συνήθως) από 3 νότες, από την χαμηλότερη στην υψηλότερη αυτές ονομάζονται:

- ρίζα ή τονική
- τρίτη [μεγάλη (+ 4 ημιτόνια) ή μικρή (+3 ημιτόνια)]
- πέμπτη [ελαττωμένη(+6 ημιτόνια),καθαρή(+7 ημιτόνια) ή αυξημένη(+8 ημιτόνια)]

Η αρμονική συσχέτιση αυτή μπορεί να παρασταθεί γεωμετρικά από ένα άπειρο δίκτυο που ονομάζεται tonnetz (Euler,1739) όπου οι κοντινότεροι κόμβοι αναπαριστούν αρμονική συσχέτιση και διαφορετικές κατασκευές στο grid αναπαριστούν διαφορετικές συγχορδίες.

Παίρνοντας 12 κόμβους του tonnetz (12 ημιτόνια απο το C έως B) και αναπαριστώντας τους σε ένα 6-διάστατο χώρο μας δίνεται η δυνατότητα να μοντελοποιήσουμε ένα συνολύλευμα απο τόνους (π.χ. μία συγχορδία) ως ένα 6-διάστατο σημείο, το οποίο ονομάζεται τονικό κεντροειδές [17]. Καθώς δε μπορούμε να οπτικοποιήσουμε ένα 6-

Figure 2.14: Παράδειγμα ισχύος τονικότητας, τονικότητας και καθαρότητας τονικότητας (από [26])

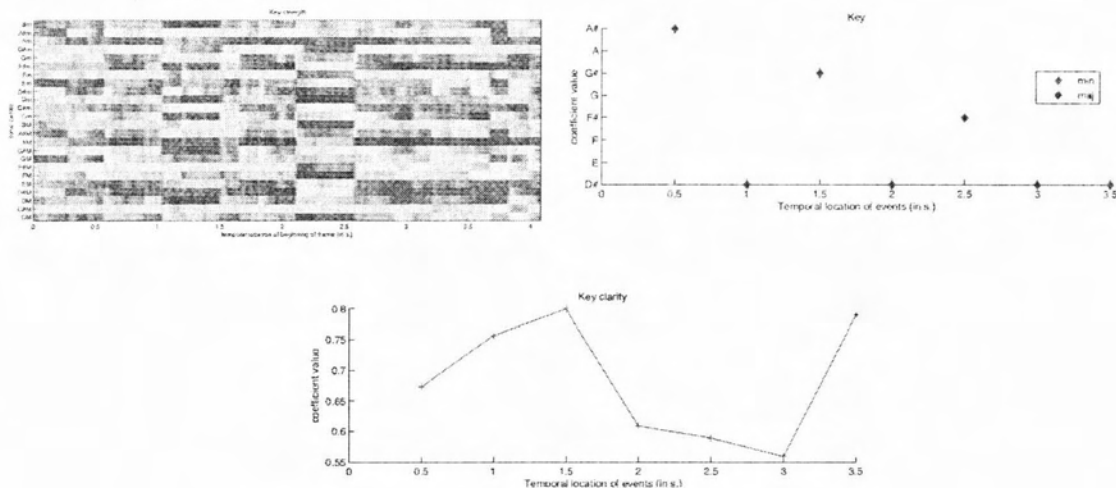
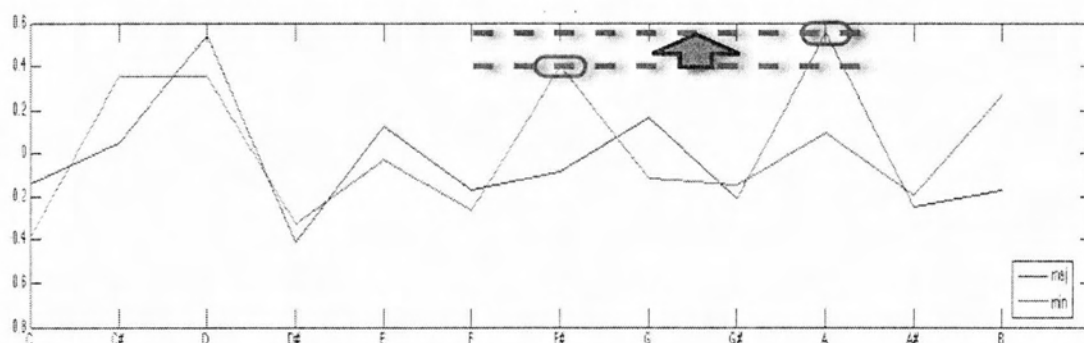


Figure 2.15: Παράδειγμα υπολογισμού τρόπου (από [26])



διάστατο χώρο, είναι χρήσιμο να φανταστούμε τρεις δισδιάστατους κύκλους συντεταγμένων όπου κάθε ένας αντιστοιχεί στις τονικές αποστάσεις πέμπτης, τρίτης μεγάλης και τρίτης μικρής (βλέπε Σχήμα 2.17).

Για τον υπολογισμό του τονικού κεντροειδούς, σαν πρώτο βήμα, υπολογίζουμε το χρωμόγραμμα c της στιγμής που αναλύουμε και στη συνέχεια το πολλαπλασιάζουμε με τον πίνακα μετασχηματισμού Φ , κανονικοποιώντας ως προς την L1 νόρμα του c :

$$\zeta_n(d) = \frac{1}{\|c_n\|} \sum_{l=0}^{11} \Phi(d, l) c_n(l),$$

$$\text{με } \begin{matrix} 0 \leq d \leq 5 \\ 0 \leq l \leq 11 \end{matrix},$$

Figure 2.16: Tonnetz Harmonic Network (από [17])

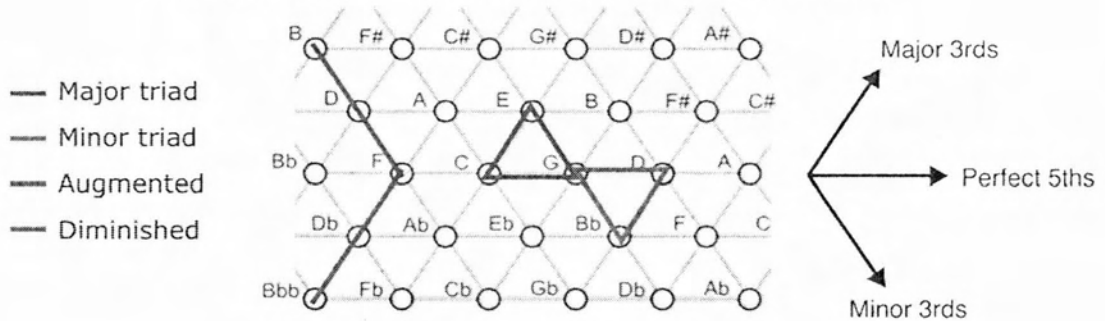
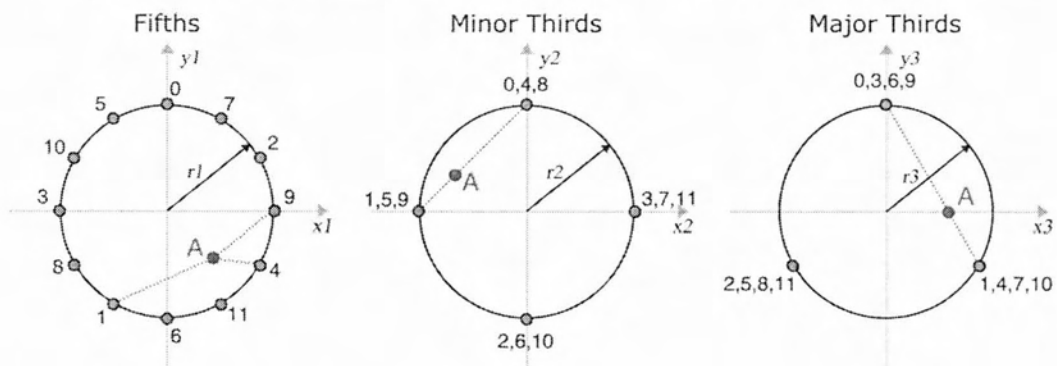


Figure 2.17: Οπτικοποίηση του 6-D τονικού χώρου ως τρεις κύκλοι. Από αριστερά προς τα δεξιά: Πέμπτης, Τρίτης μικρής, Τρίτης μεγάλης. Το σημείο A είναι τονικό κεντροειδές της συγχορδίας Λα Ματζόρε (ημιτόνια 9,1,4) (από [17])



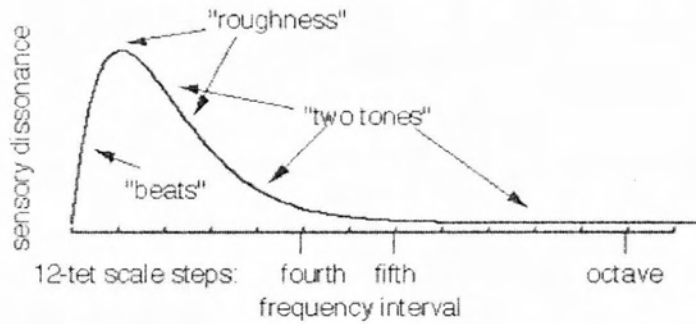
όπου d , η διάσταση που υπολογίζουμε, l το διάνυσμα χρώματος που χρησιμοποιείται και $\Phi = [\varphi_0, \varphi_1, \dots, \varphi_{11}]$,

όπου

$$\varphi_l = \begin{bmatrix} \Phi(0, l) \\ \Phi(1, l) \\ \Phi(2, l) \\ \Phi(3, l) \\ \Phi(4, l) \\ \Phi(5, l) \end{bmatrix} = \begin{bmatrix} r_1 \sin l \frac{7\pi}{6} \\ r_1 \cos l \frac{7\pi}{6} \\ r_2 \sin l \frac{3\pi}{2} \\ r_2 \cos l \frac{3\pi}{2} \\ r_3 \sin l \frac{2\pi}{3} \\ r_3 \cos l \frac{2\pi}{3} \end{bmatrix}$$

όπου $0 \leq l \leq 11$.

Figure 2.18: Μοντέλο τραχύτητας Plomp Levelt (από [26])



Οι τιμές των r_1 , r_2 και r_3 αντιπροσωπεύουν τις ακτίνες των τριών κύκλων του Σχήματος 2.17 και έχουν τις τιμές 1, 1 και 0.5 αντίστοιχα, δίνοντας μικρότερη τιμή στον κύκλο της τρίτης μικρής.

Το τονικό κεντροειδές μπορεί να χρησιμοποιηθεί για να εντοπίσουμε αλλαγές στην αρμονία στη διάρκεια ενός frame. Το HCDF ξ_n (Harmonic Change Detection function) είναι ο ρυθμός αλλαγής ενός τονικού κεντροειδούς. Συγκεκριμένα, είναι η ευκλείδεια απόσταση μεταξύ των διανυσμάτων ζ_{n-1} και ζ_{n+1} :

$$\xi_n = \sqrt{\sum_{d=0}^5 [\zeta_{n+1}(d) - \zeta_{n-1}(d)]^2}$$

Για παράθυρο 1s και επικάλυψη 50% υπολογίστηκε το mode, το key clarity και το hcdf. Ο μέσος όρος τους και η τυπική απόκλιση του mode και του key clarity, καθώς και ο μέσος όρος του hcdf επιλέχθηκαν ως χαρακτηριστικά (82-86).

2.6.4 Τραχύτητα

Η τραχύτητα (Roughness / sensory dissonance) ενός ήχου είναι η αίσθηση διακύμανσης της έντασης λόγω συγχρούσεων ηχητικών κυμάτων που είτε είναι δημιουργικές (η φάση του ενός προσεγγίζει τη φάση του άλλου), είτε καταστροφικές (οι φάσεις τους έχουν διαφορά κοντά στις 180°). Οι Plomp and Levelt (1965) πρότειναν μία προσέγγιση της τραχύτητας σε σχέση με την απόσταση των συχνοτήτων δύο ημιτονοειδών κυμάτων (βλέπε Σχήμα 2.18).

Μία από τις δημοφιλέστερες εξισώσεις για τον υπολογισμό της τραχύτητας δύο ημιτονοειδών συχνοτήτων f_1 και f_2 είναι η παρακάτω (Sethares, 1998) :

$$R = e^{-b_1 s(f_2 - f_1)} - e^{-b_2 s(f_2 - f_1)},$$

όπου $b_1 = 3.5$, $b_2 = 5.75$ και $s = x/(s_1 f_1 + s_2)$,

με $x = 0.24$, $s_1 = 0.0207$, $s_2 = 18.96$.

Υπολογίζουμε την τραχύτητα ενός δείγματος χρησιμοποιώντας όπου f_1 και f_2 τον μέσο όρο όλων των πιθανών ζευγαριών peaks απο την ανίχνευση peaks του φάσματος.

2.6.5 Δυσαρμονία

Απόηχοι (overtones) ονομάζονται οι υψηλότερες, από την κεντρική, συχνότητες που παράγονται λόγω ταλαντώσεων. Όταν αυτές οι συχνότητες είναι ακέραια πολλαπλάσια της κεντρικής συχνότητας ονομάζονται αρμονικές. Όργανα που παράγουν αρμονικές είναι το πιάνο, το βιολί ή (σε πολύ αισθητό βαθμό) το ινδικό σιτάρ, ενώ όργανα που παρεκκλίνουν οι απόηχοι τους από τις αρμονικές είναι τα κρουστά ή τα χάλκινα. Η δυσαρμονία (inharmonicities) ενός ήχου είναι ο βαθμός που οι απόηχοι που παράγονται αποκλίνουν από τις αρμονικές. Σαν χαρακτηριστικό, η δυσαρμονία μετράται σε κλίμακα από το 0 έως το 1 και υπολογίζεται ως το ποσοστό της ενέργειας που βρίσκεται εκτός των ακέραιων παραγώγων της κυρίαρχης συχνότητας. Η κυρίαρχη συχνότητα υπολογίζεται από επιλογή του peak της καμπύλης αυτοσυσχέτισης του φάσματος.

Για μέγεθος παραθύρου 46ms και 50% hop υπολογίστηκε η τραχύτητα και η δυσαρμονία. Ο μέσος όρος τους και η τυπική απόκλιση του inharmonicity είναι τα χαρακτηριστικά 87-89.

Table 2.1: Διανύσμα Χαρακτηριστικών

ΚΑΤΗΓΟΡΙΑ	NO.	ΧΑΡΑΚΤΗΡΙΣΤΙΚΟ
Mfcc's	1-13	MFCC mean
	14-26	MFCC std deviation
	27-39	MFCC delta mean
	40-52	MFCC delta std deviation
Δυναμικά	53	RMS mean
	54	RMS std deviation
	55	Attack time mean
	56	Attack time std deviation
	57	Attack slope mean
	58	Attack slope std deviation
	59	Low energy mean
Ρυθμικά	60	Tempo change mean
	61	Tempo change std deviation
	62	Metrical Centroid mean
	63	Metrical Centroid std deviation
	64	Metrical Strength mean
	65	Metrical Strength std deviation
	66	Pulse clarity
	67	Event density
Φασματικά	68	Zero-cross mean
	69	Zero-cross std deviation
	70	Rolloff85 mean
	71	Rolloff85 std deviation
	72	Centroid mean
	73	Centroid std deviation
	74	Spread mean
	75	Skewness mean
	76	Entropy mean
	77	Entropy std deviation
	78	Flux mean
	79	Flatness mean
Αρμονικά	80	Irregularity mean
	81	Irregularity std deviation
	82	Mode mean (majorness)
	83	Mode std deviation
	84	Key clarity mean
	85	Key clarity std deviation
	86	Hcdf mean
	87	Roughness mean
	88	Inharmonicity mean
	89	Inharmonicity std deviation

3 Ταξινομητές

3.1 Εισαγωγή

Στα προβλήματα αναγνώρισης προτύπων σημαντικό ρόλο παίζει η χρήση του κατάλληλου ταξινομητή, δηλαδή του αλγορίθμου που επιλέγει σε ποιά κατηγορία (κλάση) ανήκει κάθε ένα από τα δείγματα. Σε αυτό το κεφάλαιο θα γίνει συνοπτική αναφορά στις σημαντικότερες κατευθύνσεις ταξινόμησης στο πεδίο της ανάκτησης μουσικής πληροφορίας και στη συνέχεια αναλυτική περιγραφή των ταξινομητών που χρησιμοποιήθηκαν στα πειράματα (Gaussian Mixture Model και Support Vector Machine).

3.2 Κατηγορίες Ταξινόμησης σε Συστήματα Αναγνώρισης Μουσικού Συναισθήματος

Στο πεδίο της αναγνώρισης συναισθήματος σε μουσική (Music Emotion Recognition - MER) έχουν χρησιμοποιηθεί ποικίλες στρατηγικές κατηγοριοποίησης με διαφορετικά θετικά και αρνητικά η κάθε μια.

Η δημοφιλέστερη προσέγγιση ταξινόμησης λόγω της επιτυχίας και της απλότητάς της είναι η κατηγορηματική ταξινόμηση σε έναν αριθμό κλάσεων συναισθημάτων (π.χ. χαρά, ηρεμία, φόβος, λύπη). Σε αυτά τα συστήματα οι ταξινομητές που δείχνουν να λειτουργούν καλύτερα είναι οι k-Nearest Neighbour (k-NN), Support Vector Machine (SVM) και Gaussian Mixture Model (GMM)[?]. Το σημαντικότερο μειονέκτημα της κατηγορηματικής προσέγγισης είναι η δυσκολία της ανάθεσης ενός κομματιού σε μία μόνο κλάση συναισθήματος καθώς η μουσική προκαλεί μια πολυδιάστατη εμπειρία συναισθημάτων. Ως λύση σε αυτό το πρόβλημα έχουν προταθεί διάφορες προσεγγίσεις, όπως ταξινόμηση σε περισσότερες της μίας κλάσης (multi-label classification), ασαφής ταξινόμηση (fuzzy classification) και παλινδρόμηση συναισθήματος (emotion regression).

Στην ταξινόμηση σε πολλές κλάσεις (multi-label classification) τα παραδείγματα εκπαίδευσης ανατίθενται σε περισσότερες της μίας κλάσης και στη συνέχεια γίνεται η εκπαίδευσή τους με κάποιο ταξινομητή πολλαπλών ετικετών (multi-label classifier). Ο ταξινομητής με τα καλύτερα αποτελέσματα, σύμφωνα με την βιβλιογραφία, είναι ο Calibrated Label Ranking SVM (CLR_{SVM})[?].

Στα συστήματα ασαφούς ταξινόμησης (fuzzy classification) γίνεται κατηγοριοποίηση των παραδειγμάτων σε ασαφή διανύσματα (fuzzy vectors), τα οποία δείχνουν τη σχετική

ισχύ των συναισθημάτων σε έναν αριθμό από κλάσεις. Για παράδειγμα, ένα ασαφές διάνυσμα $[0.1, 0.0, 0.8, 0.1]$ υποδεικνύει σχετική ισχύ για το συναίσθημα που αντιπροσωπεύει η τρίτη κλάση. Οι δημοφιλέστεροι ταξινομητές ασαφούς λογικής είναι οι Fuzzy k-Nearest Neighbour (FkNN) και Fuzzy Nearest Mean (FNM) [?].

Τέλος, πολύ δημοφιλής προσέγγιση είναι η παλινδρόμηση σε συνεχή μοντέλα συναισθημάτων (emotion regression). Χρησιμοποιώντας έναν συνεχή χώρο, όπως το διδιάστατο χώρο διέγερσης/σθένους (arousal/valence) Thayer-Russell [33], μπορεί να γίνει περιγραφή ενός μουσικού δείγματος εκπαίδευσης ως σημείο. Σε αυτή την προσέγγιση στόχος είναι να ελαχιστοποιήσουμε το σφάλμα μεταξύ της πραγματικής τιμής από την τιμή που παράγει ο παλινδρομητής (regressor), δεδομένου ενός διανύσματος χαρακτηριστικών. Διάφοροι αλγόριθμοι έχουν δοκιμαστεί, εκ των οποίων καλύτερα αποτελέσματα πετυχαίνει ο Support Vector Regression (SVR) [?].

Σε αυτήν την εργασία εφαρμόστηκε κατηγορηματική προσέγγιση, κυρίως λόγω της ευκολίας σύνθεσης μεγάλης βάσης δεδομένων. Δοκιμάστηκαν οι ταξινομητές Gaussian Mixture Model (GMM) και Support Vector Machine (SVM).

3.3 Μοντέλο Μείγματος Γκαουσιανών Κατανομών

Το μοντέλο μείγματος Γκαουσιανών κατανομών (Gaussian Mixture Model - GMM) είναι μία παραμετρική συνάρτηση πυκνότητας πιθανότητας που παράγεται από γραμμικό συνδυασμό Γκαουσιανών κατανομών. Χρησιμοποιείται συχνά για μοντελοποίηση κατανομών πιθανοτήτων σε χαρακτηριστικά βιομετρικών συστημάτων όπως τα φασματικά χαρακτηριστικά φωνής σε ένα σύστημα αναγνώρισης ομιλητή. Το GMM δίνεται από την παρακάτω συνάρτηση:

$$p(x|\lambda) = \sum_{i=1}^M w_i g(x|\mu_i, \Sigma_i),$$

όπου x είναι ένα D -διάστατο διάνυσμα δεδομένων (π.χ. ένα διάνυσμα χαρακτηριστικών), $g(x|\mu_i, \Sigma_i)$, $i = 1, \dots, M$ είναι οι Γκαουσιανές κατανομές D διάστασης που συνδυάζονται γραμμικά και $\lambda = \{w_i, \mu_i, \Sigma_i\}$ οι παράμετροι του μοντέλου, όπου w_i , $i = 1, \dots, M$, τα βάρη των μειγμάτων, μ_i το διάνυσμα μέσου όρου και Σ_i ο πίνακας συνδιασποράς τους.

Δεδομένου μίας σύνθεσης GMM και ενός διανύσματος χαρακτηριστικών εκπαίδευσης επιθυμούμε να υπολογίσουμε το λ , δηλαδή τις παραμέτρους που προσαρμόζουν καλύτερα τα χαρακτηριστικά στην κατανομή. Αυτό επιτυγχάνεται μέσω της μεθόδου εκτίμησης μέγιστης πιθανοφάνειας (Maximum Likelihood Estimation), η οποία υπολογίζει τις παραμέτρους που μεγιστοποιούν την πιθανότητα, δεδομένου ενός διανύσματος εκπαίδευσης. Ο δημοφιλέστερος αλγόριθμος εκτίμησης της μέγιστης πιθανοφάνειας είναι

ο επαναληπτικός αλγόριθμος EM (Expectation-Maximization). Ξεκινώντας με ένα αρχικό μοντέλο λ (συνήθως υπολογίζεται μέσω του αλγορίθμου k-means), ο EM υπολογίζει ένα μοντέλο λ' , έτσι ώστε $p(X|\lambda') \geq p(X|\lambda)$. Στη συνέχεια, αρχικοποιεί το καινούργιο μοντέλο ως λ και η διαδικασία επαναλαμβάνεται έως ένα επιθυμητό όριο σύγκλισης.

Μέσω της παραπάνω διαδικασίας μπορούμε να υπολογίσουμε τις κατανομές των μειγμάτων των Γκαουσιανών που ταιριάζουν καλύτερα στα χαρακτηριστικά εκπαίδευσης της κάθε κλάσης. Έχοντας, λοιπόν, τις συναρτήσεις πυκνότητας πιθανότητας κάθε κλάσης, εύκολα μπορούμε να αποφανθούμε την κλάση στην οποία πιθανοτικά ανήκει ένα καινούργιο διάνυσμα χαρακτηριστικών. Φυσικά, σημαντικός παράγοντας για τον υπολογισμό του καταλληλότερου Γκαουσιανού μείγματος μίας κλάσης είναι και ο βέλτιστος αριθμός M των μειγμάτων που θα χρησιμοποιηθούν. Στα πειράματα της παρούσας εργασίας μετά από σύντομες δοκιμές ο αριθμός των μειγμάτων σταθεροποιήθηκε στο 1. Επίσης, χρησιμοποιήθηκε διαγώνιος πίνακας συνδιασποράς, καθώς θεωρήθηκε πως τα χαρακτηριστικά είναι ασυσχέτιστα.

3.4 Μηχανές Διανυσματικής Στήριξης

Οι μηχανές διανυσματικής στήριξης (Support Vector Machines - SVM) είναι ταξινομητές μέγιστου περιθωρίου (maximal margin classifiers) και χρησιμοποιούνται στην πλειοψηφία των προβλημάτων εξαγωγής μουσικής πληροφορίας. Πρόκειται για ένα σετ αλγορίθμων που απεικονίζουν διανύσματα χαρακτηριστικών x σε υπερεπίπεδα $\phi(x)$, με σκοπό τα διανύσματα που ανήκουν στην ίδια κατηγορία να απεικονιστούν στο ίδιο υπερεπίπεδο. Συγκεκριμένα, στην γραμμική μορφή της, μια μηχανή διανυσματικής στήριξης εφαρμόζεται σε χώρους με διανύσματα 2 κατηγοριών και υπολογίζει μία διαχωριστική ευθεία ώστε πρώτον, τα διανύσματα ίδιας κλάσης να ανήκουν στο ίδιο υπερεπίπεδο και δεύτερον, η απόσταση της από τα κοντινότερα παραδείγματα (και των 2 κατηγοριών) να είναι μέγιστη. Τα παραδείγματα αυτά καλούνται διανύσματα στήριξης (Support Vectors) και η μέγιστη απόσταση τους από την ευθεία που διαχωρίζει τα υπερεπίπεδα ονομάζεται περιθώριο (Margin).

Η χρήση των SVM μπορεί να γενικευτεί και σε προβλήματα όπου τα χαρακτηριστικά δεν είναι γραμμικά διαχωρίσιμα χρησιμοποιώντας το λεγόμενο κόλπο πυρήνα (kernel trick). Δεδομένου ενός συνόλου χαρακτηριστικών x_1, x_2, \dots, x_n , πυρήνας (kernel) είναι μια συνάρτηση ομοιότητας $f_i = k(x_i, x_n)$ μεταξύ ενός χαρακτηριστικού x_i με όλα τα x_n δυνατά σημεία. Χρησιμοποιώντας τα f_i αντί για τα διανύσματα x_i στην εξίσωση της γραμμικής μηχανής διανυσματικής στήριξης υπολογίζουμε ένα μη γραμμικό τρόπο να διαχωρίσουμε τις κλάσεις σε υπερεπίπεδα. Η χρήση κατάλληλου πυρήνα οδηγεί σε καλύτερο διαχωρισμό των κλάσεων, ανάλογα με το πρόβλημα. Οι πιο δημοφιλείς πυρήνες για SVM είναι:

- Πολυωνυμικός (Polynomial):

$$k(x_i, x_n) = (x_i x_n + m)^d,$$

όπου d, m ακέραιοι.

- Ακτινικής Συνάρτησης Βάσης (Radial Basis Function):

$$k(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2),$$

για $\gamma > 0$, όπου $\gamma = 1/2\sigma^2$.

Ο SVM είναι ένας ταξινομητής δύο κλάσεων, ο οποίος όμως μπορεί να χρησιμοποιηθεί σε προβλήματα ταξινόμησης πολλαπλών κλάσεων. Υπάρχουν, κυρίως, δύο προσεγγίσεις για την επιλογή της κυρίαρχης κλάσης:

- Μια εναντίων μίας (One vs one): Σε αυτήν την περίπτωση εκπαιδεύουμε τόσους δυαδικούς SVM όσες και οι κλάσεις. Αν, για παράδειγμα, έχουμε τις κλάσεις A, B, C, D τότε εκπαιδεύουμε τέσσερις δυαδικούς SVM: A εναντίων όχι A, B εναντίων όχι B, C εναντίων όχι C και D εναντίων όχι D. Στην συνέχεια επιλέγουμε την θετική κλάση η οποία είναι σε μεγαλύτερη απόσταση από το περιθώριο (margin).
- Μία εναντίων όλων (One vs all): Σε αυτήν την περίπτωση γίνεται εκπαίδευση όλων των δυνατών ζευγών ταξινόμησης. Επιλέγεται η κλάση με τη μέγιστη επιτυχία.

4 Επιλογή Χαρακτηριστικών

4.1 Εισαγωγή

Η επιλογή ενός υποσυνόλου από το διάνυσμα χαρακτηριστικών είναι μία πολύ σημαντική διαδικασία στα προβλήματα ταξινόμησης. Υπάρχουν πολλοί λόγοι για να μειώσουμε τον αριθμό των χαρακτηριστικών καταλήγοντας σε ένα επαρκώς ελάχιστο υποσύνολο. Η υπολογιστική πολυπλοκότητα είναι ο προφανής λόγος. Ένας ακόμα λόγος είναι ότι πολλά από τα χαρακτηριστικά που έχουν επιλεγεί είναι πιθανό να εκφράζουν την ίδια πληροφορία, οπότε η υπολογιστική πολυπλοκότητα αυξάνεται χωρίς ιδιαίτερο λόγο. Σε μία χειρότερη περίπτωση κάποια από τα χαρακτηριστικά που επιλέγονται είναι θορυβώδη, με αποτέλεσμα όχι μόνο να αυξάνεται ο χρόνος εκτέλεσης αλλά και να μειώνεται η επιτυχία ταξινόμησης. Σκοπός αυτού του κεφαλαίου είναι πρώτον, η συνοπτική αναφορά μεθόδων επιλογής χαρακτηριστικών και δεύτερον, η ανάλυση των μεθόδων που δοκιμάστηκαν στα πειράματα της παρούσας εργασίας (Stepwise Selection και Random Mutation Hill Climbing).

4.2 Μέθοδοι Επιλογής Χαρακτηριστικών

Οι μέθοδοι επιλογής χαρακτηριστικών (Feature Selection Methods) είναι συνδυασμός τεχνικών αναζήτησης υποσυνόλων και μεθόδων αξιολόγησής τους. Ο πιο απλός αλγόριθμος είναι να δοκιμάσουμε όλους τους πιθανούς συνδυασμούς υποσυνόλων με σκοπό να βρούμε το υποσύνολο το οποίο πετυχαίνει καλύτερα αποτελέσματα. Βέβαια, αυτός ο αλγόριθμος δεν μπορεί να χρησιμοποιηθεί σε πραγματικά προβλήματα με δεκάδες ή εκατοντάδες χαρακτηριστικά, καθώς τα δυνατά υποσύνολα αυξάνονται εκθετικά με τον αριθμό των χαρακτηριστικών. Οπότε, σαν ρεαλιστική λύση στο πρόβλημα της επιλογής χαρακτηριστικών επιλέγουμε συνήθως μία εκ των δύο προσεγγίσεων:

- Προσέγγιση Συνολικότητας (Wrapper Approach): Οι μέθοδοι συνολικότητας (wrapper) χρησιμοποιούν ένα μοντέλο πρόβλεψης για να βαθμολογήσουν ένα υποσύνολο. Πρόκειται για αλγορίθμους που εκπαιδεύουν ένα υποσύνολο χαρακτηριστικών και βαθμολογούν την απόδοση του, συγκρίνοντας την επιτυχία του σε σχέση με κάποιο υποσύνολο προηγούμενου σταδίου. Καθώς οι μέθοδοι αυτοί απαιτούν την εκπαίδευση του κάθε καινούργιου υποσυνόλου, είναι υπολογιστικά ακριβοί, ιδιαίτερα για μεγάλο σύνολο χαρακτηριστικών. Επίσης τα αποτελέσματά τους ανταποκρίνονται στην απόδοση του συγκεκριμένου ταξινομητή, άρα δεν

γενικεύονται για κάθε είδους ταξινόμηση. Συνήθεις μέθοδοι συνολικότητας είναι: Εξαντλητική (Exhaustive), Σταδιακής Επιλογής (Stepwise Selection), Προσομοίωσης Ανόπτησης (Simulated Annealing), Γενετική (Genetic) κ.ά.

- Προσέγγιση Φίλτρου (Filter Approach): Οι μέθοδοι φίλτρου χρησιμοποιούν κάποια μετρική υπολογισμού της απόδοσης του υποσυνόλου, χωρίς να γίνει η δοκιμή του στο σύστημα ταξινόμησης. Συγκεκριμένα για κάθε συνδυασμό χαρακτηριστικών χρησιμοποιείται κάποιο κριτήριο διαχωρισμού και επιλέγεται ο καλύτερος συνδυασμός. Συνηθισμένα κριτήρια διαχωρισμού είναι τα Mutual Information, Pearson product-moment correlation coefficient, και inter/intra class distance. Τα φίλτρα είναι λιγότερο υπολογιστικά ακριβά από της μεθόδους συνολικότητας, αλλά για συγκεκριμένα συστήματα ταξινόμησης πετυχαίνουν, συνήθως, μικρότερη απόδοση.

4.3 Μέθοδοι Σταδιακής Επιλογής (Stepwise Selection)

Οι μέθοδοι σταδιακής επιλογής (stepwise selection) είναι υποβέλτιστες μέθοδοι επιλογής χαρακτηριστικών που χρησιμοποιούνται συχνά λόγω της απλότητας υλοποίησής τους. Οι δύο πιο συνηθισμένες μέθοδοι είναι:

- Οπισθοδρομική Απαλοιφή Χαρακτηριστικών (Backward Feature Elimination - BFE): Σε αυτόν τον αλγόριθμο ξεκινάμε με ένα σύνολο $x = \{x_1, x_2, \dots, x_m\}$, όπου m ο αριθμός των χαρακτηριστικών. Υπολογίζουμε όλα τα δυνατά υποσύνολα $m - 1$ διάστασης και επιλέγουμε το υποσύνολο x' το οποίο πετυχαίνει καλύτερη απόδοση στην ταξινόμηση. Στη συνέχεια, επαναλαμβάνουμε τη διαδικασία θέτοντας $x = x'$ έως ότου γίνει απαλοιφή όλων των χαρακτηριστικών. Θεωρούμε ως βέλτιστο υποσύνολο το υποσύνολο το οποίο πέτυχε καλύτερα αποτελέσματα στην διάρκεια του αλγορίθμου.
- Εμπρόσθια Επιλογή Χαρακτηριστικών (Forward Feature Selection - FFE): Πρόκειται για την αντίστροφη διαδικασία από την προηγούμενη. Ξεκινώντας, δηλαδή, με το κενό υποσύνολο $x = \{\}$, σε κάθε επανάληψη επιλέγουμε το χαρακτηριστικό το οποίο, εάν προστεθεί στο τρέχων σύνολο, μας δίνει μεγαλύτερη ακρίβεια. Ο αλγόριθμος τερματίζει όταν όλα τα χαρακτηριστικά προστεθούν στο σύνολο x . Επιλέγεται το υποσύνολο με την καλύτερη επιτυχία στην διάρκεια της εκτέλεσής του αλγορίθμου.

4.4 Μέθοδος Αναρρίχησης Λόφου Τυχαίας Μετάλλαξης (Random Mutation Hill Climbing)

Οι δύο προηγούμενοι αλγόριθμοι υποφέρουν από τη λεγόμενη επίδραση εμφωλιασμού (nesting effect). Δηλαδή από τη στιγμή που ένα χαρακτηριστικό απορριφθεί κατά τη

4.4 Μέθοδος Αναρρίχησης Λόφου Τυχαίας Μετάλλαξης (Random Mutation Hill Climbing)

διάρκεια της οπισθοδρομικής απαλοιφής δεν υπάρχει η δυνατότητα να επανεξεταστεί. Το αντίθετο συμβαίνει στη διάρκεια της εμπρόσθιας επιλογής, δηλαδή από τη στιγμή που ένα χαρακτηριστικό επιλέγεται δεν μπορεί να απορριφθεί αργότερα. Σε αυτήν την εργασία επιλέχθηκε ο αλγόριθμος αναρρίχησης λόφου τυχαίας μετάλλαξης (Random Mutation Hill Climbing) ως λύση σε αυτό το πρόβλημα. Πρόκειται για έναν ευριστικό αλγόριθμο συνολικότητας που χρησιμοποιεί την τυχειότητα για να προσεγγίσει το βέλτιστο υποσύνολο. Συγκεκριμένα, μοντελοποιεί το σύνολο των χαρακτηριστικών σε ένα δυαδικό διάνυσμα (επιλογή / όχι επιλογή) και προσπαθεί να προσεγγίσει το βέλτιστο υποσύνολο αλλάζοντας κάθε φορά ένα τυχαίο bit. Σε πρώτη φάση ο αλγόριθμος επιλέγει ένα τυχαίο διάνυσμα, και σε κάθε επανάληψη “μεταλλάσσει” (αντιστρέφει) ένα τυχαίο bit. Εάν το καινούριο υποσύνολο δίνει καλύτερα αποτελέσματα κατά την ταξινόμηση από το τρέχον υποσύνολο, η αλλαγή παραμένει. Εάν όχι, το μεταλλαγμένο bit αντιστρέφεται ξανά. Ο αλγόριθμος τερματίζεται μετά από έναν αριθμό επαναλήψεων t_{max} και επιλέγεται το καλύτερο υποσύνολο.

Ο παραπάνω αλγόριθμος είναι πιθανό να λύσει το πρόβλημα της εσωτερικής συσχέτισης μεταξύ των χαρακτηριστικών, αλλά παρουσιάζει άλλα σημαντικά μειονεκτήματα. Πρώτον, λόγω της ευριστικής του φύσης δεν μπορούμε να γνωρίζουμε εάν το επιλεγμένο υποσύνολο είναι βέλτιστο και δεύτερον, λόγω της άπληστης συμπεριφοράς του αλγορίθμου δεν αποκλείεται να υπολογίσουμε ένα τοπικό μέγιστο το οποίο δεν μπορεί να ξεπεραστεί αυξάνοντας τον αριθμό των επαναλήψεων. Τέλος, καθώς σε έναν υπολογιστή δεν μπορούμε να παράγουμε πραγματικά τυχαίους αριθμούς, η επιτυχία ενός αλγορίθμου που βασίζεται στην τυχειότητα εξαρτάται από την ποιότητα της γεννήτριας τυχαίων αριθμών που χρησιμοποιούμε.

5 Συλλογή της Βάσης Δεδομένων και Προεπεξεργασία

5.1 Εισαγωγή

Λόγω έλλειψης μιας κοινής βάσης δεδομένων μουσικών κομματιών-συναισθημάτων, οι περισσότεροι ερευνητές οδηγούνται στο να συνθέσουν δίκες τους βάσεις. Η επιθυμητή βάση θα πρέπει να καλύπτει όλο το φάσμα των ειδών της μουσικής και να προσπαθεί να εξαλείψει φαινόμενα επανάληψης του ίδιου καλλιτέχνη (artist effects). Πρέπει, επίσης, να αντιπροσωπεύει την πλειοψηφία της άποψης του μέσου ακροατή, όσον αφορά την ετικέτα συναισθήματος που θα χρησιμοποιηθεί. Ένας από τους πιο συνηθισμένους τρόπους σύνθεσης βάσης γενικής αλήθειας (Ground-Truth Data Collection) είναι η συλλογή δεδομένων από χειρονακτικό σχολιασμό. Δηλαδή, συλλογή συναισθηματικών αντιδράσεων από ομάδες ατόμων κατά τη διάρκεια της μουσικής ακρόασης. Η παραπάνω μέθοδος, όμως, είναι πολύ “ακριβή” όσον αφορά το ανθρώπινο δυναμικό που απαιτείται και συνήθως καταλήγει σε δημιουργία μικρών βάσεων με λιγότερα από 1000 κομμάτια. Μία ακόμα δυσκολία είναι η εύρεση στατιστικού δείγματος, ώστε να αντιπροσωπεύει το σύνολο της μουσικής κουλτούρας σε παγκόσμιο επίπεδο. Σε αντίθεση με την παραπάνω μέθοδο, στο πρώτο μέρος του κεφαλαίου προτείνουμε τη σύνθεση βάσης δεδομένων μέσω άντλησης πληροφορίας από το διαδίκτυο. Στο δεύτερο μέρος του κεφαλαίου, θα περιγράψουμε την διαδικασία προεπεξεργασίας της.

5.2 Η Συλλογή της Βάσης Δεδομένων

Με την ανάπτυξη της μουσικής ανακάλυψης μέσω διαδικτύου και την επιτυχία διαδικτυακών ραδιοφώνων και κοινωνικών δικτύων με επίκεντρο τη μουσική (Last.FM, Spotify, Streameoood [3, 8, 9] κ.ά.) είναι εφικτή η άντληση μεγάλου όγκου μουσικής πληροφορίας. Μπορούμε να εξάγουμε συναισθηματική πληροφορία μέσω των “ετικετών” που δίνουν οι χρήστες σε κάθε κομμάτι (user tags) με αποτέλεσμα τη δημιουργία μεγαλύτερων και πιο αντικειμενικών βάσεων δεδομένων. Συγκεκριμένα, για την συλλογή της βάσης δεδομένων που χρησιμοποιήθηκε στα πειράματα ακολουθήθηκε η παρακάτω διαδικασία:

Πρώτα επιλέχθηκε η πηγή για τη συλλογή πληροφορίας. Χρησιμοποιήθηκε το βρετανικό site Last.FM[3], πρώτον ως ένα από τα πιο δημοφιλή μουσικά portal αυτή τη στιγμή και δεύτερον λόγω της δυνατότητας δωρεάν χρήσης της διεπαφής προγραμματισμού

εφαρμογών (Application Programming Interface - API) που διαθέτει. Για την άντληση πληροφορίας από το API του Last.FM χρησιμοποιήθηκε η βιβλιοθήκη pylast.py για python [6].

Σαν πρώτο βήμα έγινε η συλλογή όλων των ετικετών χρηστών (user tags) από το Last.FM μέσω της διεπαφής pylast.py για python και στη συνέχεια έγινε απαλοιφή όσων ετικετών είχαν λιγότερες από 100 παρουσίες. Έπειτα αφαιρέθηκαν από τη λίστα όλα τα είδη μουσικής σύμφωνα με τη λίστα μουσικών ειδών του Wikipedia [4]. Από τα εναπομείναντα στοιχεία επιλέχθηκαν όσα είχαν συναισθηματικό περιεχόμενο σύμφωνα με λεξικά συναισθημάτων και διάθεσης [20, 13], αλλά και την αντίστοιχη εργασία του Xiao Hu [19]. Για τις παραπάνω διαδικασίες χρησιμοποιήθηκαν εντολές του Unix Shell. Σε αυτό το σημείο χρησιμοποιήθηκε χειρονακτική απαλοιφή ορισμένων αμφισβητήσιμων στοιχείων (π.χ. το συναισθηματικό disturbed αντιστοιχεί στο όνομα πολύ δημοφιλούς συγκροτήματος) καταλήγοντας σε συνολικά 102 ετικέτες. Στη συνέχεια, έγινε σταδιακή ομαδοποίηση των στοιχείων σε κοινές κατηγορίες συνωνύμων ή παρόμοιων συναισθημάτων από δύο άτομα αρίστους γνώστες της αγγλικής γλώσσας. Οι κατηγορίες που συγκεντρώθηκαν εμφανίζονται στον Πίνακα 5.1.

Από αυτές τις κατηγορίες επιλεχτήκαν οι τέσσερις πιο ακραίες μεταξύ τους σύμφωνα με το μοντέλο του Russell[33]. Δηλαδή τα συναισθήματα που αντιστοιχούν στα τέσσερα τεταρτημόρια του διδιάστατου επιπέδου σθένους-έντασης (valence-arousal). Συγκεκριμένα, επιλέχθηκαν οι κατηγορίες “Χαρά”, “Λύπη”, “Ηρεμία”, “Θυμός” (“Happy”, “Sad”, “Calm”, “Angry”). Για κάθε ετικέτα από τις τέσσερις κατηγορίες έγινε η άντληση των τίτλων των κομματιών με τις περισσότερες ψήφους ανά ετικέτα (μέσω του API του Last.FM). Αντλήσαμε τα 50 κορυφαία κομμάτια ανά ετικέτα με εξαίρεση τις ετικέτες της κατηγορίας “Θυμός”, όπου λόγω των λίγων συνωνύμων που εντοπίστηκαν αναγκαστήκαμε να πάρουμε τα 200 κορυφαία κομμάτια κάθε ετικέτας (ώστε ανά κατηγορία να έχω περίπου ίσο αριθμό τραγουδιών, γύρω στα 1500). Τέλος, υλοποιήθηκε ένας κώδικας python για αναζήτηση και τοπικό κατέβασμα δειγμάτων μορφής mp3 (30 ή 60 δευτερολέπτων). Σε αυτόν τον κώδικα χρησιμοποιήθηκε το API της αμερικάνικης σελίδας 7digital [2] η οποία περιέχει χιλιάδες samples από δημοφιλή τραγούδια. Η επιλογή από τα αποτελέσματα της αναζήτησης έγινε με βάση τη μικρότερη απόσταση Levenshtein από τη συμβολοσειρά αναζήτησης. Τέλος, έγινε χειρονακτική απαλοιφή των δειγμάτων που δεν αντιπροσώπευαν τον τίτλο που αναζητήθηκε. Δυστυχώς δεν ήταν εφικτή η εύρεση μουσικών δειγμάτων όλων των κομματιών, αλλά και πάλι το μέγεθος της βάσης δεδομένων (4023 κομμάτια) ξεπερνά σε αριθμό κομματιών τις περισσότερες βάσεις δεδομένων της βιβλιογραφίας. Το αποτέλεσμα της συλλογής της βάσης φαίνεται στον Πίνακα 5.2.

5.3 Προεπεξεργασία των Δεδομένων (Data Pre-Processing)

Η προεπεξεργασία των δεδομένων και η οργάνωση της βάσης σε τμήματα εκπαίδευσης και δοκιμής είναι απαραίτητη διαδικασία σε κάθε πρόβλημα ταξινόμησης.

Σαν πρώτο βήμα, έγινε μετατροπή των 4023 δειγμάτων, 30 και 60 δευτερολέπτων, από κωδικοποίηση mp3 σε standard format wav. Συγκεκριμένα, χρησιμοποιήθηκε το πρόγραμμα sox με το οποίο έγινε μετατροπή των δειγμάτων από στερεοφωνικό mp3 ακρίβειας 24 bit και συχνότητας δειγματοληψίας 44100 Hz, σε μορφή wav, με ακρίβεια 16 bit, ενός καναλιού και συχνότητα δειγματοληψίας 22050 Hz.

Στη συνέχεια, έγινε διαχωρισμός της βάσης σε τρία τμήματα (σε τρεις καταλόγους): εκπαίδευσης, δοκιμής και επικύρωσης. Χρησιμοποιώντας Unix Shell επιλέχθηκε τυχαία το 70% της βάσης για εκπαίδευση (training set), το 15% για δοκιμή (test set) και το υπόλοιπο 15% για επαλήθευση (validation set).

Συγκεκριμένα,

- Το τμήμα εκπαίδευσης (training set) περιέχει συνολικά 2762 δείγματα: 721 της κατηγορίας “Λύπη” (“Sad”), 614 της κατηγορίας “Χαρά” (“Happy”), 720 της κατηγορίας “Ηρεμία” (“Calm”) και 707 δείγματα της κατηγορίας “Θυμός” (“Anger”).
- Το τμήμα δοκιμής (test set) περιέχει συνολικά 593 δείγματα: 155 της κατηγορίας “Λύπη” (“Sad”), 132 της κατηγορίας “Χαρά” (“Happy”), 154 της κατηγορίας “Ηρεμία” (“Calm”) και 152 δείγματα της κατηγορίας “Θυμός” (“Anger”).
- Το τμήμα επικύρωσης (validation set) περιέχει συνολικά 593 δείγματα: 155 της κατηγορίας “Λύπη” (“Sad”), 132 της κατηγορίας “Χαρά” (“Happy”), 154 της κατηγορίας “Ηρεμία” (“Calm”) και 152 δείγματα της κατηγορίας “Θυμός” (“Anger”).

Τέλος, έγινε μετονομασία των δειγμάτων για λόγους οργάνωσης. Σε κάθε δείγμα ανά κατάλογο ανατέθηκε μία ετικέτα της κατηγορίας στο οποίο ανήκει και ένας μοναδικός αριθμός δείγματος. Για παράδειγμα, ang432.wav, hap145.wav, sad344.wav, cal219.wav κ.λ.π.

Επίσης, επιλέχτηκαν τυχαία δύο κομμάτια από κάθε κατηγορία (συνολικά 10 κομμάτια) για τη δημιουργία ενός υποσυνόλου ασφαλείας (test subset), για δοκιμές και επικύρωση ορθότητας κατά τη συγγραφή του κώδικα εξαγωγής χαρακτηριστικών και ταξινόμησης.

Table 5.1: Κατηγορίες που προέκυψαν και παραδείγματα ετικετών

Κατηγορίες Παρόμοιων Συναισθημάτων	Παραδείγματα Ετικετών
CALM	mellow,relaxing,calm. . .
SAD	sad,melancholy,heartbreak. . .
HAPPY	happy,feelgood,positive..
ANGRY	angry,aggressive,rage..
ROMANTIC	romantic,love songs,love song...
ENERGETIC	energetic
SENSUAL	sensual,sex
ANGST	angst,angsty
SLEEP	sleep,music to fall asleep to...
INSPIRATIONAL	inspirational,inspiring,wistfull
DREAMY	dream, dreamy
DARK	gloomy,darkness
HAPPY&SAD	bittersweet,happysad
PLAYFUL	playful
DRAMATIC	dramatic
FUNNY	funny,lol,humor
HAUNTING	haunting
COMFORT	comfort music,comfort
CYNICAL	cynical,sarcastic..
EXCITING	exciting,thrilling. . .

Table 5.2: Συνολικός αριθμός δειγμάτων mp3 ανά κατηγορία

Κατηγορία	Αριθμός Δειγμάτων
Angry	1031 samples
Calm	1041 samples
Happy	889 samples
Sad	1062 samples
Σύνολο	4023 samples

6 Πειράματική Διαδικασία και Αποτελέσματα

Στα προηγούμενα κεφάλαια έγινε περιγραφή των χαρακτηριστικών, των ταξινομητών και αλγορίθμων επιλογής χαρακτηριστικών που χρησιμοποιήθηκαν. Σε αυτό το κεφάλαιο αναλύουμε την πειραματική διαδικασία και παραθέτουμε τα αποτελέσματα που συγκεντρώθηκαν. Όλες οι συναρτήσεις που χρησιμοποιήθηκαν για εξαγωγή χαρακτηριστικών ανήκουν στην βιβλιοθήκη `mirtoolbox` (v1.5) για `matlab` [27]. Όλα τα πειράματα έτρεξαν σε υπολογιστή με διπύρνηνο επεξεργαστή AMD E450 1.65 GHz και 4 GB μνήμη.

6.1 Πείραμα 1: Εκπαίδευση GMM με Χαρακτηριστικά MFCC

Σε αυτό το αρχικό πείραμα δοκιμάστηκε η επιτυχία των χαρακτηριστικών MFCC σε εκπαίδευση Μοντέλου Γκαουσιανών Μειγμάτων (GMM).

6.1.1 Εξαγωγή των MFCC

Αρχικά έγινε εξαγωγή των 13 πρώτων συντελεστών MFCC των δειγμάτων της βάσης δεδομένων. Χρησιμοποιήθηκε η συνάρτηση `mirframe()` για παραθύρωση των κυματομορφών σε παράθυρα των 50ms και επικάλυψη 50% και η συνάρτηση `mirmfcc()` για εξαγωγή των MFCC. Στην συνέχεια, υπολογίστηκε η πρώτη παράγωγος αυτών μέσω της συνάρτησης `diff()` του Matlab (delta MFCC). Τέλος, υπολογίσαμε τον μέσο όρο και την τυπική απόκλιση των MFCC και delta MFCC, ώστε να καταλήξουμε σε διάνυσμα 4×13 (52) διάστασης.

6.1.2 Εκπαίδευση των GMM

Χρησιμοποιώντας τις συναρτήσεις `gmminit()` και `gmmem()` της βιβλιοθήκης `netlab` [32] για Matlab, δημιουργήσαμε μία συνάρτηση `gmm_classify()` η οποία, δεδομένου των συνόλων εκπαίδευσης και δοκιμής, επιστρέφει την επιτυχία ταξινόμησης κάθε δείγματος δοκιμής (accuracy) και τον πίνακα σύγχυσης (confusion matrix).

Table 6.1: Αποτελέσματα Εκπαίδευσης GMM με χαρακτηριστικά MFCC (όπου μ ο μέσος όρος, σ η τυπική απόκλιση και Δ η παράγωγος). Με έντονη γραμμοτοσείρα εμφανίζεται η μέγιστη επιτυχία ανα κατηγορία.

Χαρακτηριστικά	Θυμός	Ηρεμία	Χαρά	Λύπη	Μέσος Όρος
μ MFCC	73.02%	40.90%	48.48%	24.51%	46.73%
μ, σ MFCC	77.63%	44.80%	39.39%	29.03%	47.71%
$\mu, \Delta\mu$ MFCC	73.68%	42.20%	40.90%	21.93%	44.68%
$\mu, \sigma, \Delta\mu, \Delta\sigma$ MFCC	71.71%	31.16%	48.48%	54.83%	51.55%

Επιτυχία Ταξινόμησης (%) του Σετ Δοκιμής.

Χαρακτηριστικά	Θυμός	Ηρεμία	Χαρά	Λύπη	Μέσος Όρος
μ MFCC	71.05%	53.89%	47.72%	27.09%	49.94%
μ, σ MFCC	72.36%	44.15%	46.21%	18.70%	45.36%
$\mu, \Delta\mu$ MFCC	71.71%	51.94%	43.18%	28.37%	48.80%
$\mu, \sigma, \Delta\mu, \Delta\sigma$ MFCC	69.73%	38.31%	58.33%	51.61%	54.49%

Επιτυχία Ταξινόμησης (%) του Σετ Επικύρωσης.

Στην συνέχεια, δοκιμάστηκε η συνάρτηση `gmm_classify` με το διάνυσμα των MFCC για κάθε δείγμα του συνόλου δοκιμής αλλά και επικύρωσης. Δοκιμάστηκε εκπαίδευση με διάφορους αριθμούς Γκαουσιανών με διαγώνιους ή ολόκληρους πίνακες συνδιασποράς. Τα καλύτερα αποτελέσματα δόθηκαν από μείγμα μόνο μίας Γκαουσιανής κατανομής με διαγώνιο πίνακα συνδιασποράς.

Σε πρώτη φάση έγινε εκπαίδευση με μοναδικά χαρακτηριστικά τις 13 τιμές του μέσου όρου των MFCC. Στην συνέχεια δοκιμάστηκε ο μέσος όρος και η τυπική απόκλιση τους, έπειτα ο μέσος όρος των MFCC και των πρώτων παραγώγων τους (Delta MFCC) και τέλος ο μέσος όρος και η διασπορές και των MFCC και των παραγώγων τους. Στον Πίνακα 6.1 αναγράφεται η ακρίβεια πρόβλεψης για κάθε τρέξιμο του αλγόριθμου ταξινόμησης ανά κατηγορία συναισθήματος.

6.1.3 Αποτελέσματα

Τα αποτελέσματα του Πίνακα 6.1 δείχνουν ότι στην περίπτωση που και τα 52 χαρακτηριστικά MFCC χρησιμοποιηθούν στην εκπαίδευση έχουμε την καλύτερη επιτυχία ταξινόμησης. Επίσης, μπορούμε να παρατηρήσουμε ότι το σετ δοκιμής δίνει παρόμοια αποτελέσματα με το σετ επικύρωσης, δείγμα της επιτυχίας της βάσης δεδομένων. Τέλος, παρατηρούμε ότι μεγαλύτερη ακρίβεια έχουμε στην κατηγορία “Θυμός” λόγω της μεγαλύτερης ετερογένειας των μουσικών ιδιωμάτων που την απαρτίζουν (heavy metal, punk rock κ.ά.)

6.2 Πείραμα 2: Εκπαίδευση GMM με το Σύνολο των Χαρακτηριστικών

Στο πείραμα αυτό δοκιμάστηκε η επιτυχία του ταξινομητή GMM στο σύνολο των χαρακτηριστικών αλλά και σε κάθε κατηγορία ξεχωριστά.

6.2.1 Εξαγωγή των Χαρακτηριστικών

Για την εξαγωγή των χαρακτηριστικών χρησιμοποιήθηκε εξ' ολοκλήρου η βιβλιοθήκη *mirtoolbox* [27]. Το συνολικό διάνυσμα 89 χαρακτηριστικών παρουσιάζεται στον Πίνακα 2.1. Για αναλυτική περιγραφή των χαρακτηριστικών ο αναγνώστης μπορεί να απευθυνθεί στο Κεφάλαιο 2.

Επίσης να αναφέρουμε, ότι για ευκολία κατά την εναλλαγή των χαρακτηριστικών, μοντελοποιήσαμε την χρήση, ή όχι, ενός χαρακτηριστικού ως ένα bit σε διάνυσμα απο bits. Για παράδειγμα, εάν θέλουμε να εκπαιδεύσουμε τον ταξινομητή με τα πρώτα 13 χαρακτηριστικά δηλώνουμε κατά την αρχικοποίηση:

```
...  
feature_bitmap(1:13)=1;  
...
```

6.2.2 Εκπαίδευση των GMM

Έχοντας εξάγει τον συνολικό αριθμό των χαρακτηριστικών, εκπαιδεύσαμε ένα GMM ανά κατηγορία χαρακτηριστικών. Συγκεκριμένα εκπαιδεύσαμε από ένα GMM στα χαρακτηριστικά των κατηγοριών: MFCC (χαρακτηριστικά 1:52), Δυναμικά (χαρακτηριστικά 53:59), Ρυθμικά (χαρακτηριστικά 60:67), Φασματικά (χαρακτηριστικά 68:81) και Αρμονικά (χαρακτηριστικά 82:90). Μετά απο δοκιμές, παρατηρήσαμε ότι η μία Γκαουσιανή κατανομή ανά μείγμα ευνοεί τα MFCC και τα φασματικά χαρακτηριστικά ενώ το μείγμα 3 Γκαουσιανών ευνοεί τα χαρακτηριστικά υψηλότερου επιπέδου αλλά ρίχνει την συνολική απόδοση. Σταθεροποιήσαμε, λοιπόν, τον αριθμό των Γκαουσιανών ανά μείγμα στην μία κατανομή (χρησιμοποιώντας διαγώνιο πίνακα συνδιασποράς). Το πείραμα έτρεξε με 2 σετ δοκιμής, το σετ δοκιμής (test set) και το σετ επικύρωσης (validation set). Στον Πίνακα 6.2 αναγράφεται η ακρίβεια πρόβλεψης ταξινόμησης ανά κατηγορία χαρακτηριστικών.

6.2.3 Συμπεράσματα

Στο πείραμα αυτό, έγινε δοκιμή του ταξινομητή GMM ανά κατηγορία χαρακτηριστικών. Παρατηρούμε, ότι το σύνολο όλων των χαρακτηριστικών δίνει τα καλύτερα

Table 6.2: Αποτελέσματα Εκπαίδευσης GMM μία Γκαουσιανής κατανομής ανά κατηγορία χαρακτηριστικών. Με έντονη γραμμοσειρά εμφανίζεται η μέγιστη επιτυχία ανά κατηγορία.

Χαρακτηριστικά	Θυμός	Ηρεμία	Χαρά	Λύπη	Μέσος Όρος
MFCC	71.71%	31.16%	48.48%	54.83%	51.55%
Δυναμικά	18.42%	27.27%	77.27%	40%	40.74%
Ρυθμικά	48.02%	46.10%	64.93%	23.87%	45.59%
Φασματικά	76.31%	48.05%	37.87%	26.45%	47.17%
Αρμονικά	69.07%	42.20%	50%	32.25%	48.38%
Συνολικά	73.68%	45.45%	59.09%	47.74%	56.49%

Επιτυχία Ταξινόμησης (%) του Σετ Δοκιμής.

Χαρακτηριστικά	Θυμός	Ηρεμία	Χαρά	Λύπη	Μέσος Όρος
MFCC	69.73%	38.31%	58.33%	51.61%	54.49%
Δυναμικά	17.76%	42.20%	74.24%	36.77%	42.74%
Ρυθμικά	52.63%	49.35%	60.60%	21.93%	46.13%
Φασματικά	76.31%	61.03%	38.63%	27.74%	50.93%
Αρμονικά	73.02%	48.05%	42.42%	32.25%	48.94%
Συνολικά	72.36%	57.14%	59.84%	35.48%	56.21%

Επιτυχία Ταξινόμησης(%) του Σετ Επικύρωσης.

αποτελέσματα και στα 2 σετ δοκιμής και μάλιστα με παρόμοια απόδοση. Η παρόμοια αυτή απόδοση είναι δείγμα επιτυχίας διαχωρισμού της βάσης δοκιμής σε ποιοτικά ισάξια 2 μέρη.

Επίσης, μπορούμε να βγάλουμε ενδιαφέροντα συμπεράσματα παρατηρώντας την επιτυχία κάθε κατηγορίας χαρακτηριστικών ανά κλάση συναισθήματος. Συγκεκριμένα, παρατηρούμε ότι την υψηλότερη απόδοση στην κλάση του θυμού μας δίνει η εκπαίδευση των φασματικών χαρακτηριστικών (συμπεριλαμβανομένου και των MFCC). Αυτό είναι λογικό λόγω της διαφορετικότητας της χροιάς αυτών των κομματιών (παραμορφωμένες ηλεκτρικές κιθάρες, παραμορφωμένες φωνητικές γραμμές κ.ά.). Παρόμοια έχουμε υψηλή συσχέτιση φασματικών χαρακτηριστικών στην κατηγορία της ηρεμίας, πιθανά λόγω των “ομαλών” κυματομορφών των κομματιών αυτής της κατηγορίας. Στην κατηγορία του συναισθήματος της χαράς παρατηρούμε υπεροχή των δυναμικών και ρυθμικών χαρακτηριστικών, που μπορεί να εξηγηθεί λόγω της σταθερότητας έντασης και ρυθμού που συνήθως έχει ένα “χαρούμενο” κομμάτι. Τέλος, στο συναίσθημα της λύπης έχουμε υψηλότερη απόδοση όταν εκπαιδεύουμε χαρακτηριστικά MFCC. Μία πιθανή αιτία για αυτό είναι ότι τα περισσότερα κομμάτια αυτής της κλάσης στηρίζονται στην ερμηνεία του τραγουδιστή, οπότε η κλίμακα του Mel που αποδίδει βέλτιστα στα προβλήματα ανθρωπίνης ομιλίας είναι κατάλληλη.

6.3 Πείραμα 3: Βελτίωση των αποτελεσμάτων μέσω γραμμικού συνδιασμού των Γκαουσσισιανών

Παρατηρώντας την υπεροχή συγκεκριμένων κατηγοριών χαρακτηριστικών στα διαφορετικά συναισθήματα, αποφασίσαμε να δοκιμάσουμε τον γραμμικό συνδυασμό των διαφορετικών μειγμάτων και στην συνέχεια να αναζητήσουμε τα βέλτιστα βάρη τους.

6.3.1 Πειραματική Διαδικασία

Έστω c το Γκαουσσισιανό μείγμα και Ω_i οι πίνακες χαρακτηριστικών των 5 κατηγοριών χαρακτηριστικών (MFCC, δυναμικά, ρυθμικά, φασματικά, αρμονικά). Θέλουμε να βρούμε τα βέλτιστα βάρη w_i ώστε η πιθανότητα

$$p_{max} = \sum_{i=1}^5 w_i p(c|\Omega_i)$$

του γραμμικού συνδιασμού των πιθανοτήτων ταξινόμησης $p(c|\Omega_i)$ να είναι μέγιστη.

Σε αυτό το πρόβλημα δοκιμάσαμε μια απλοϊκή προσέγγιση αναζήτησης Monte Carlo. Πρόκειται για ευριστική αναζήτηση της βέλτιστης λύσης μέσω παραγωγής τυχαίων λύσεων και αξιολόγηση τους (στην συγκεκριμένη περίπτωση με κριτήριο το μέγιστο p_{max}) επαναληπτικά. Στη συγκεκριμένη περίπτωση σε κάθε επανάληψη παρήγαμε ένα τυχαίο διάνυσμα $[w_1, w_2, w_3, w_4, w_5]$, ώστε κάθε βάρος να είναι ένας τυχαίος αριθμός από το σύνολο $\{0.000, 0.005, 0.010, 0.015, \dots, 1.00\}$ με την εξαίρεση ότι $w_5 = 1 - (w_1 + w_2 + w_3 + w_4)$, έτσι ώστε το άθροισμα των βαρών να είναι 1. Τρέχοντας τον αλγόριθμο σε επαρκή χρόνο προσεγγίσαμε το διάνυσμα των βαρών που βελτιώνουν την απόδοση κατά την ταξινόμηση του σετ δοκιμής. Στη συνέχεια, χρησιμοποιώντας τα βάρη που συνθέσαμε, έγινε δοκιμή του γραμμικού συνδιασμού στο σετ επικύρωσης. Τα αποτελέσματα του πειράματος αναγράφονται στον Πίνακα 6.3.

6.3.2 Συμπεράσματα

Στον Πίνακα 6.3 παρατηρούμε ότι τα βέλτιστα βάρη που εντοπίσαμε έχουν μικρή διακύμανση από το ισοβαρές μοντέλο με εξαίρεση το βάρος των ρυθμικών και mfcc χαρακτηριστικών, που εμφανίζουν αντίστοιχα μικρή αύξηση και μείωση. Καθώς το κέρδος ταξινόμησης του σετ δοκιμής δεν φαίνεται να επαληθεύεται από το σετ επικύρωσης, τα αποτελέσματα αυτά δεν μπορούν να γενικευτούν. Αντί της αξιολόγησης κατηγοριών χαρακτηριστικών, στα επόμενα πειράματα προτείνουμε την αναζήτηση βέλτιστου υποσυνόλου χαρακτηριστικών μέσω αλγορίθμων επιλογής.

Table 6.3: Βέλτιστα βάρη για συνδιασμό Γκαουσιανών και αποτελέσματα ταξινόμησης.

Κατηγορία Χαρακτηριστικών	Βάρος γραμμικού συνδιασμού
Mfcc	0.165
Δυναμικά	0.195
Ρυθμικά	0.240
Φασματικά	0.195
Αρμονικά	0.205

	Ίσα βάρη	Βάρος πειράματος	Ποσοστιαίο Κέρδος
Επιτυχία Δοκιμής	56.49%	58.16%	+2.96%
Επιτυχία Επικύρωσης	56.21%	55.25%	-1.71%

6.4 Πειραμα 4: Επιλογή Χαρακτηριστικών για Εκπαίδευση με GMM

Σε αυτό το πείραμα έγινε προσπάθεια αναζήτησης του καλύτερου υποσυνόλου χαρακτηριστικών. Δοκιμάστηκε ο αλγόριθμος συνολικότητας Οπισθοδρομικής Απαλοιφής (Backward Feature Elimination - BFE) και ο αλγόριθμος Αναρρίχησης Λόφου Τυχαίας Μετάλλαξης (Random Mutation Hill Climbing - RMHC). Ο αναγνώστης μπορεί να απευθυνθεί στο Κεφάλαιο 4 για αναλυτική περιγραφή των δύο αλγορίθμων.

6.4.1 Οπισθοδρομική Απαλοιφή Χαρακτηριστικών

Έγινε υλοποίηση του αλγορίθμου σε Matlab. Χρησιμοποιήσαμε το σετ δοκιμής κατά την αναζήτηση και το σετ επικύρωσης για επαλήθευση του αποτελέσματος. Το βέλτιστο υποσύνολο και η επιτυχία του υποσυνόλου αυτού κατά την ταξινόμηση αναγράφεται στον Πίνακα 6.4.

6.4.2 Αναρρίχηση Λόφου Τυχαίας Μετάλλαξης

Όπως αναφέραμε στο Κεφάλαιο 4 οι αλγόριθμοι σταδιακής επιλογής υποφέρουν από το πρόβλημα της μη επανεξέτασης του υποσυνόλου αναζήτησης, για αυτό τον λόγο υλοποιήσαμε το αλγόριθμο Αναρρίχησης Λόφου Τυχαίας Μετάλλαξης. Η υλοποίηση του έγινε σε Matlab. Χρησιμοποιήσαμε το σετ δοκιμής κατά την αναζήτηση και το σετ επικύρωσης για επαλήθευση του αποτελέσματος. Ο αλγόριθμος έτρεξε για 10000 επαναλήψεις συνολικά, με κάθε φορά που εντοπίζεται πιθανό τοπικό μέγιστο να γίνεται επαναρχικοποίηση. Το βέλτιστο υποσύνολο και η επιτυχία του υποσυνόλου αυτού κατά την ταξινόμηση αναγράφεται στον Πίνακα 6.5.

Table 6.4: Βέλτιστο υποσύνολο και επιτυχία Οπισθοδρομικής Απαλοιφής για GMM.

Υποσύνολο Χαρακτηριστικών (28/89):

MFCC	2ο, 9ο, 10ο, 11ο μ Mfcc 9ο, 10ο, 11ο σ Mfcc 1ο, 2ο μ Δ(Mfcc) 3ο, 9ο, 10ο, 12ο σ Δ(Mfcc)
Δυναμικά	μ Attack time, σ Attack time,
Ρυθμικά	μ Tempo Change, σ Tempo Change
Φασματικά	μ Zero cross, μ Centroid, μ Entropy, μ Flux, μ Irregularity, σ Irregularity
Αρμονικά	μ mode, σ mode, μ Keyclarity, μ Roughness, μ Inharmonicity

	Σύνολο Χαρακτηριστικών	Υποσύνολο O.A.X	Ποσοστιαίο Κέρδος
Επιτυχία Δοκιμής	56.49%	61.34%	+8.59%
Επιτυχία Επικύρωσης	56.21%	55.91%	-0.53%

Κέρδος Ταχύτητας Εκτέλεσης: 128% (Απο 5.7 s σε 2.5 s)

6.4.3 Συμπεράσματα

Από τα αποτελέσματα των Πινάκων 6.4 και 6.5 παρατηρούμε ότι η επιλογή υποσυνόλου χαρακτηριστικών κατά την ταξινόμηση προσφέρει ικανοποιητική απόδοση στο σύστημα μας.

Στην περίπτωση της επιλογής με οπισθοδρόμηση παρατηρούμε ότι με το 30% των χαρακτηριστικών πετυχαίνουμε πολύ καλύτερη απόδοση στο σετ δοκιμής άλλα μικρή μείωση στο σετ επικύρωσης. Παρόλο που δεν επαληθεύτηκε η υπεροχή του υποσυνόλου από το συνολικό διάνυσμα των χαρακτηριστικών, μπορούμε να θεωρήσουμε την απόδοση του ικανοποιητική, λόγω της μειωμένης υπολογιστικής πολυπλοκότητας που προσφέρει.

Στην περίπτωση του αλγορίθμου RMHC, παρατηρούμε σημαντική βελτίωση της απόδοσης στο σετ δοκιμής άλλα και μικρή βελτίωση στο σετ επικύρωσης. Βέβαια τα αποτελέσματα επικύρωσης δεν προσφέρουν αρκετή σταθερότητα ώστε να θεωρήσουμε το υποσύνολο βέλτιστο. Επίσης καθώς περιέχει μόνο το 40% των χαρακτηριστικών, προσφέρει μειωμένη υπολογιστική πολυπλοκότητα.

Σαν συμπέρασμα από τα παραπάνω δεδομένα, μπορούμε να διαπιστώσουμε ότι υπάρχει επικάλυψη στο μεγαλύτερο μέρος της πληροφορίας στο σύνολο των χαρακτηριστικών. Αυτό είναι φανερό λόγω των πολύ καλών αποτελεσμάτων με πολύ μικρότερο σετ χαρακτηριστικών. Επίσης καθώς σκοπός των αλγορίθμων ήταν η καλύτερη απόδοση του

Table 6.5: Βέλτιστο υποσύνολο και επιτυχία RMHC για GMM.

Υποσύνολο Χαρακτηριστικών (38/89):

MFCC	1ο,2ο,4ο,6ο,7ο,8ο,9ο,11ο,13ο μ Mfcc 3ο,12ο,13ο σ Mfcc 1ο,5ο,7ο,11ο μ Δ (Mfcc) 1ο,2ο,6ο,7ο,9ο,12ο σ Δ (Mfcc)
Δυναμικά	σ Attack time, μ Attack slope σ Attack slope, μ Lowenergy Ratio
Ρυθμικά	μ Tempo Change, σ Tempo Change σ Metroid Strength, EventDensity
Φασματικά	μ Entropy, μ Irregularity
Αρμονικά	μ mode, σ mode, μ Keyclarity, σ Keyclarity μ Roughness, μ Inharmonicity

	Σύνολο Χαρακτηριστικών	Υποσύνολο Α.Α.Τ.Μ.	Ποσοστιαίο Κέρδος
Επιτυχία Δοκιμής	56.49%	61.74%	+9.29%
Επιτυχία Επικύρωσης	56.21%	56.44%	+0.44%

Κέρδος Ταχύτητας Εκτέλεσης: 84% (Απο 5.7 s σε 3.1 s)

υποβέλτιστου υποσυνόλου, μπορούμε να πούμε ότι ο RMHC απέδωσε καλύτερα από την οπισθοδρομική απαλοιφή, δεδομένου όμως μεγαλύτερου χρόνου εκτέλεσης.

6.5 Πειραμα 5: Εκπαίδευση με SVM

6.5.1 Υλοποίηση του SVM

Σύμφωνα με την βιβλιογραφία, ο ταξινομητής με την καλύτερη απόδοση στα συστήματα αναγνώρισης συναισθήματος σε μουσική είναι οι μηχανές διανυσματικής στήριξης (Support Vector Machines). Η υλοποίηση έγινε χρησιμοποιώντας τις συναρτήσεις svmtrain και svmpredict της βιβλιοθήκης libsvm [12]. Έγινε χρήση του πυρήνα (kernel) RBF, λόγω της υπεροχής του στα αποτελέσματα των πειραμάτων. Είναι σημαντικό να αναφέρουμε, επίσης, ότι έγινε αρχική κανονικοποίηση των δεδομένων. Η κανονικοποίηση είναι απαραίτητη για την χρήση πυρήνα RBF, καθώς η συνάρτηση ομοιότητας περιλαμβάνει την Ευκλείδεια απόσταση μεταξύ των δεδομένων στο επίπεδο των χαρακτηριστικών. Χρησιμοποιήθηκε η συνάρτηση κανονικοποίησης zscore (z score normalization) η οποία υπολογίζει την κανονικοποιημένη τιμή ως τον λόγο της διαφοράς της τιμής με το μέσο όρο των τιμών, προς την τυπική απόκλιση της. Δηλαδή, $z = (x - \mu) / \sigma$, όπου μ , ο μέσος όρος και σ , η τυπική απόκλιση. Τέλος, να αναφέρουμε ότι ακολουθήθηκε

Table 6.6: Αποτελέσματα Εκπαίδευσης SVM.

Θυμός	Ηρεμία	Χαρά	Λύπη	Μέσος Όρος
76.82%	51.95%	59.85%	46.15%	58.52%

Επιτυχία Ταξινόμησης (%) του Σετ Δοκιμής.

Θυμός	Ηρεμία	Χαρά	Λύπη	Μέσος Όρος
78.15%	56.49%	62.12%	46.15%	60.54%

Επιτυχία Ταξινόμησης (%) του Σετ Επικύρωσης.

η προσέγγιση one-vs-one (default προσέγγιση της βιβλιοθήκης libsvm) για την ταξινόμηση των 4 κατηγοριών συναισθημάτων.

6.5.2 Πειραματική Διαδικασία

Έχοντας εξάγει το σύνολο των χαρακτηριστικών από τα προηγούμενα πειράματα, έγινε ταξινόμηση του συνολικού διανύσματος χαρακτηριστικών με χρήση του SVM. Το πείραμα έτρεξε με 2 σετ δοκιμής, το σετ δοκιμής (test set) και το σετ επικύρωσης (validation set). Στον Πίνακα 6.6 αναγράφεται η ακρίβεια πρόβλεψης ταξινόμησης του SVM.

6.5.3 Συμπεράσματα

Απο τα αποτελέσματα του Πίνακα 6.6 βλέπουμε ότι ο SVM σαν ταξινομητής μας δίνει καλύτερα συνολικά αποτελέσματα από τον GMM. Η παραπάνω ακρίβεια επιβεβαιώνεται και από το σετ επικύρωσης

6.6 Πείραμα 6: Επιλογή Χαρακτηριστικών για Εκπαίδευση με SVM

Έχοντας επαληθεύσει την επιτυχία των αλγορίθμων επιλογής για τον ταξινομητή GMM, σε αυτό το πείραμα δοκιμάζουμε την επιλογή χαρακτηριστικών μέσω μεθόδων συνολικότητας (wrappers) στον ταξινομητή SVM. Δοκιμάστηκε ο αλγόριθμος Εμπρόσθιας Σταδιακής Επιλογής (Forward Feature Selection - FFS) και ο αλγόριθμος Αναρρίχησης Λόφου Τυχαίας Μετάλλαξης (Random Mutation Hill Climbing - RMHC).

6.6.1 Εμπρόσθια Επιλογή Χαρακτηριστικών

Έγινε υλοποίηση του αλγορίθμου σε Matlab. Χρησιμοποιήσαμε το σετ δοκιμής κατά την αναζήτηση και το σετ επικύρωσης για επαλήθευση του αποτελέσματος. Το βέλτισ-

Table 6.7: Βέλτιστο υποσύνολο και επιτυχία FFS για SVM.

Υποσύνολο Χαρακτηριστικών (65/89):

MFCC	1ο,7ο,8ο,9ο,10ο,12ο,13ο μ Mfcc 1ο,3ο,5ο,6ο,7ο,8ο,9ο,10ο,12ο,13ο σ Mfcc 1ο,2ο,3ο,4ο,6ο,7ο,9ο,10ο,11ο,12ο μ Δ (Mfcc) 1ο-13ο σ Δ (Mfcc)
Δυναμικά	μ RMS Energy, σ RMS Energy, σ Attack time, σ Attack time μ Attack slope, σ Attack slope, μ Lowenergy Ratio
Ρυθμικά	μ Tempo Change, Metroid Clarity μ , Metroid Clarity σ , μ Metroid Strength, EventDensity
Φασματικά	μ Zero cross, σ Zero cross, μ Rolloff85, μ Skewness μ Entropy, σ Entropy, μ Flux, μ Flatness
Αρμονικά	μ mode, σ mode, μ Keyclarity, μ HCDF μ Roughness

	Σύνολο Χαρακτηριστικών	Υποσύνολο Ε.Ε.Χ	Ποσοστιαίο Κέρδος
Επιτυχία Δοκιμής	56.49%	64.25%	+13.74%
Επιτυχία Επικύρωσης	56.21%	62.39%	+10.99%

Κέρδος Ταχύτητας Εκτέλεσης: 14% (Απο 8.1 s σε 7.1 s)

το υποσύνολο και η επιτυχία του υποσυνόλου αυτού κατά την ταξινόμηση αναγράφεται στον Πίνακα 6.7

6.6.2 Αναρρίχηση Λόφου Τυχαίας Μετάλλαξης

Όπως αναφέραμε και στην επιλογή χαρακτηριστικών για GMM, ο αλγόριθμος RMHC αποδείχτηκε πολύ καλή επιλογή για αναζήτηση υποβέλτιστου υποσυνόλου. Ο αλγόριθμος έτρεξε για 10000 επαναλήψεις με επαναρχικοποίηση σε εντοπισμό πιθανού τοπικού μέγιστου. Τα αποτελέσματα αναγράφονται στον Πίνακα 6.8

6.6.3 Συμπεράσματα

Από τα αποτελέσματα της επιλογής χαρακτηριστικών, παρατηρούμε την σημαντική βελτίωση στα αποτελέσματα κατά την επιλογή υποσυνόλου χαρακτηριστικών. Είναι σημαντικό να παρατηρήσουμε ότι το σετ επικύρωσης επιβεβαιώνει και στις 2 περιπτώσεις τα υποσύνολο χαρακτηριστικών που παρήχθησαν από το σετ δοκιμής. Στην περίπτωση της εμπρόσθιας επιλογής χρησιμοποιούμε το 73% των χαρακτηριστικών και έχουμε μέση ακρίβεια επιτυχίας 63.32%, δηλαδή περίπου 7 ποσοστιαίες μονάδες παραπάνω από

Table 6.8: Βέλτιστο υποσύνολο και επιτυχία RMHC για SVM.

Υποσύνολο Χαρακτηριστικών (46/89):

MFCC	1ο,4ο,7ο,8ο,9ο,10ο,13ο μ Mfcc 2ο,6ο,7ο,8ο,9ο,10ο,13ο σ Mfcc 3ο,4ο,7ο,10ο,11ο μ Δ (Mfcc) 1ο,2ο,3ο,4ο,5ο,6ο,8ο,11ο,12ο σ Δ (Mfcc)
Δυναμικά	μ RMS Energy, σ RMS Energy, σ Attack time σ Attack time, μ Attack slope, μ Lowenergy Ratio
Ρυθμικά	Metroid Clarity μ , EventDensity
Φασματικά	σ Zero cross, μ Rolloff85, μ Skewness, μ Entropy μ Flatness, μ Irregularity, σ Irregularity,
Αρμονικά	μ mode, μ Keyclarity, μ Inharmonicity

	Σύνολο Χαρακτηριστικών	Υποσύνολο A.A.T.M.	Ποσοστιαίο Κέρδος
Επιτυχία Δοκιμής	56.49%	64.08%	+13.44%
Επιτυχία Επικύρωσης	56.21%	61.05%	+8.61%

Κέρδος Ταχύτητας Εκτέλεσης: 59% (Απο 8.1 s σε 5.1 s)

την χρήση του συνόλου των χαρακτηριστικών. Στην περίπτωση του RMHC έχουμε μέση ακρίβεια επιτυχίας 62.56% αλλά κερδίζουμε σε ταχύτητα καθώς χρησιμοποιείται το 55% των χαρακτηριστικών. Επίσης, παρατηρούμε ότι τα 2 υποσύνολα που παρήχθησαν έχουν σημαντική ομοιότητα μεταξύ τους. Συγκεκριμένα τα 41 από τα 46 χαρακτηριστικά του υποσυνόλου του RMHC, ανήκουν στο υποσύνολο του FFS. Τέλος, μπορούμε να βγάλουμε συμπέρασμα ότι ο FFS είχε καλύτερη απόδοση από τον RMHC και μάλιστα σε μικρότερο χρόνο εκτέλεσης.

7 Συμπεράσματα

7.1 Συνολικά Συμπεράσματα των Πειραμάτων

Στην παρούσα εργασία, δοκιμάσαμε την επιτυχία των ταξινομητών Γκαουσιανου Μείγματος (Gaussian Mixture Model - GMM) και Μηχανών Διανυσματικής Στήριξης (Support Vector Machines - SVM) σε Βάση Δεδομένων που συνθέσαμε. Για εξαγωγή των χαρακτηριστικών χρησιμοποιήσαμε την βιβλιοθήκη mirtoolbox [27] και τέλος δοκιμάσαμε μεθόδους συνολικότητας για επιλογή του βέλτιστου υποσυνόλου χαρακτηριστικών.

Η μέγιστη μέση ακρίβεια (μέσος όρος του σετ δοκιμής και του σετ επικύρωσης) που πετύχαμε είναι 63.32% με χρήση 65 χαρακτηριστικών και SVM με Πυρήνα Ακτινικής Συνάρτησης Βάσης (RBF). Ο SVM απέδωσε, συνολικά, καλύτερα από τον GMM. Βέβαια, τα πολύ καλά αποτελέσματα με λίγα χαρακτηριστικά και η ταχύτητα του δεύτερου τον κάνουν μια ανταγωνιστική επιλογή.

Οι αλγόριθμοι Οπισθοδρομικής Απαλοιφής και Εμπρόσθιας Επιλογής, έδειξαν να λειτουργούν ικανοποιητικά στο πρόβλημα της επιλογής χαρακτηριστικών, όπως και ο αλγόριθμος Αναρρίχησης Λόφου Τυχαίας Μετάλλαξης, δεδομένου μεγάλου χρόνου εκτέλεσης. Εν τέλει, οι αλγόριθμοι σταδιακής επιλογής μπορούν να προσεγγίσουν πάντα μια ικανοποιητική λύση, ενώ για έναν αλγόριθμο τυχειότητας όπως τον RMHC δεν μπορούμε να είμαστε σίγουροι. Από την άλλη, δεδομένης υψηλής επεξεργαστικής ισχύς και αρκετού χρόνου είναι πιθανότερο να βρούμε τη βέλτιστη λύση χρησιμοποιώντας έναν αλγόριθμο τυχειότητας.

7.2 Συμβολή της Διπλωματικής Εργασίας

Σημαντική συνεισφορά της συγκεκριμένης διπλωματικής είναι η σύνθεση της βάσης δεδομένων δημοφιλών τραγουδιών-συναισθήματος. Στον τομέα της αναγνώρισης μουσικού συναισθήματος (Music Emotion Recognition - MER) υπάρχει μικρός αριθμός βάσεων δεδομένων, οι οποίες μάλιστα είναι σχετικά μικρού μεγέθους ώστε να θεωρήσουμε τα αποτελέσματά τους ικανοποιητικά. Η βάση γενικής αλήθειας που συνθέσαμε έχει τον μεγαλύτερο όγκο μουσικών δειγμάτων (4023 δείγματα) από κάθε άλλη βάση που γνωρίζουμε, και λαμβάνει υπόψη την πληθώρα των συναισθηματικών ετικετών που χρησιμοποιούν οι χρήστες.

Επίσης, τα αποτελέσματα της ταξινόμησης είναι πολύ ελπιδοφόρα για την ποιότητα της βάσης αλλά και την ποιότητα των χαρακτηριστικών που εξάγαμε, καθώς η επιτυχία 63.32% είναι πολύ κοντά στο κατώφλι 66% που έχει επιτευχθεί σε αντίστοιχες έρευνες [21]. Οι μέθοδοι επιλογής συνολικότητας που χρησιμοποιήθηκαν δείχνουν ότι μικρά σετ χαρακτηριστικών μπορούν να αποδώσουν πολύ ικανοποιητικά, δείγμα πως πολλά από τα χαρακτηριστικά που χρησιμοποιούνται επικαλύπτονται και πολλές φορές δίνουν θορυβώδη πληροφορία. Στην παρούσα εργασία δεν μπορούμε να υποθέσουμε ένα βέλτιστο υποσύνολο χαρακτηριστικών, αλλά παρατηρώντας τα αποτελέσματα μπορούμε να εκτιμήσουμε ότι:

1. Πάνω από το 50% της πληροφορίας δίνεται από τα χαρακτηριστικά MFCC, και
2. Έχουμε σημαντική βελτίωση με χρήση των χαρακτηριστικών: Attack time, Spectral Entropy, Mode και Key Clarity.

Ακόμα, μπορούμε να καταλήξουμε ότι ο ταξινομητής GMM είναι πλήρως ανταγωνιστικός στο πρόβλημα της αναγνώρισης μουσικού συναισθήματος και σε συστήματα που μας απασχολεί η ταχύτητα εκτίμησης θα ήταν μία πολύ καλή επιλογή.

Επίσης, συγκρίνοντας την αναζήτηση βέλτιστων βαρών για GMM και τις μεθόδους συνολικότητας για τον ίδιο ταξινομητή, μπορούμε να συμπεράνουμε ότι οι δεύτερες μας δίνουν καλύτερα αποτελέσματα στο πρόβλημα της βελτίωσης απόδοσης ενός συστήματος MER.

7.3 Μελλοντικές Ερευνητικές Κατευθύνσεις

Η κατεύθυνση που ακολουθήσαμε σε αυτήν την διπλωματική εργασία είχε άξονες την κατανόηση του τρόπου εξαγωγής χαρακτηριστικών και των αλγορίθμων ταξινόμησης και επιλογής χαρακτηριστικών. Παρόλο, που τα αποτελέσματα ήταν πλήρως ικανοποιητικά υπάρχουν διάφορες κατευθύνσεις που μπορούν να ακολουθηθούν για βελτίωση των αποτελεσμάτων αλλά και για περαιτέρω εξερεύνηση του προβλήματος της αναγνώρισης μουσικού συναισθήματος.

Η κατηγορηματική προσέγγιση, δηλαδή η ταξινόμηση σε σταθερές κλάσεις συναισθημάτων, είναι μία προσέγγιση που δεν λαμβάνει υπόψη την υποκειμενικότητα και ασάφεια του προβλήματος. Για παράδειγμα, το ότι ένα μουσικό δείγμα ταξινομήθηκε στην κλάση “Ηρεμία” δεν σημαίνει ότι δεν μπορεί να ανήκει στις κλάσεις “Χαράς” ή “Λύπης”. Σαν μελλοντική κατεύθυνση θα μπορούσε, λοιπόν, να γίνει προσπάθεια ταξινόμησης μέσω ασαφούς λογικής (fuzzy logic) ή μέσω παλινδρόμησης (regression) των δεδομένων σε συνεχή χώρο.

Ένας άλλος άξονας συνέχισης της εργασίας θα μπορούσε να είναι η δοκιμή περαιτέρω ταξινομητών. Ο αλγόριθμος k-NN για παράδειγμα έχει δείξει να λειτουργεί πολύ αποδοτικά στο συγκεκριμένο πρόβλημα. Η χρήση επιπλέον ταξινομητών θα μπορούσε να εξάγει συμπέρασμα για το ποιος ταξινομητής είναι βέλτιστος, συγκεκριμένα για το πρόβλημα αναγνώρισης συναισθήματος στην δημοφιλή σύγχρονη μουσική.

Επίσης, θα μπορούσε να δοκιμαστεί η εξαγωγή επιπλέον χαρακτηριστικών, που πιθανά παρέχουν χρήσιμη συναισθηματική πληροφορία της μουσικής. Ένα παράδειγμα που έχει δείξει την ραγδαία αύξηση της επιτυχίας ταξινόμησης στο πρόβλημα, είναι η χρήση πληροφορίας από τους στίχους του κομματιού. Αυτό είναι λογικό καθώς οι περισσότεροι ακροατές δίνουν ίση βαρύτητα στους στίχους και στη μουσική ενός κομματιού και η συναισθηματική τους εμπειρία είναι συνονθύλευμα από τους δυο αυτούς παράγοντες.

Μία ακόμα κατεύθυνση είναι η ανάπτυξη πιο περίπλοκων αλγορίθμων τυχαιότητας για επιλογή χαρακτηριστικών στο πρόβλημά μας. Η κατασκευή γενετικού αλγόριθμου αναζήτησης ή επιλογή με προσομοίωση ανόπτησης είναι δύο αλγόριθμοι που πιθανά να έδιναν καλύτερα αποτελέσματα επιλογής υποσυνόλου. Επίσης θα μπορούσε να γίνει προσπάθεια βελτιστοποίησης του κώδικα για γρηγορότερα αποτελέσματα. Ένα παράδειγμα είναι η παράλληλη υλοποίηση του RMHC.

Τέλος, χρησιμοποιώντας το σύστημα ταξινόμησης της εργασίας θα μπορούσαμε να υλοποιήσουμε μια ολοκληρωμένη εφαρμογή αυτόματης αναγνώρισης ή αυτόματης δημιουργίας λιστών αναπαραγωγής μουσικής για χρήση σε Η/Υ ή φορητές συσκευές.

Ευχαριστίες

Με την ολοκλήρωση της διπλωματικής εργασίας θα ήθελα καταρχήν να ευχαριστήσω τον κ. Γεράσιμο Ποταμιάνο για την καθοδήγησή του, την στήριξη και τις συμβουλές που μου προσέφερε σε όλη την διάρκεια της παρούσας εργασίας. Επίσης, ευχαριστώ την Μυρτώ Ζάββου για την πολύτιμες μουσικές της γνώσεις και την βοήθεια της σε όλη την διάρκεια της διπλωματικής. Τέλος ευχαριστώ την οικογένεια και τους φίλους μου που μου στάθηκαν σε όλη την διάρκεια των προπτυχιακών μου σπουδών.

Bibliography

- [1] (2012) Million song dataset. [Online]. Available: <http://labrosa.ee.columbia.edu/millionsong/>
- [2] (2013) 7digital. [Online]. Available: <http://www.7digital.com/>
- [3] (2013) Last.fm. [Online]. Available: <http://www.last.fm/>
- [4] (2013) List of music styles. [Online]. Available: http://en.wikipedia.org/wiki/List_of_music_styles
- [5] (2013) Musicoverly. [Online]. Available: <http://musicoverly.com/>
- [6] (2013) Pylast. [Online]. Available: <https://code.google.com/p/pylast/>
- [7] (2014) Kasetophono. [Online]. Available: <http://www.kasetophono.com/>
- [8] (2014) Spotify. [Online]. Available: <http://spotify.com/>
- [9] (2014) Stereomood - music for every mood. [Online]. Available: <https://www.stereomood.com/>
- [10] M. Barthet, G. Fazekas, and M. Sandler, "Multidisciplinary perspectives on music emotion recognition: Implications for content and context-based models," *Proc. CMMR*.
- [11] K. Bischoff, C. Firan, R. Paiu, W. Nejdl, C. Laurier, and M. Sordo, "Music mood and theme classification - a hybrid approach," *Proc. 10th Int. Conf. Music Inf. Retrieval ISMIR09*, 2009.
- [12] C.-C. Chang and C.-J. Lin, "Libsvm: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, p. 27, 2011.
- [13] T. Drummond. (2012) Emotion vocabulary. [Online]. Available: <http://tomdrummond.squarespace.com/emotion-vocabulary/>
- [14] M. Farmer, S. Bapna, and A. Jain, "Large scale feature selection using modified random mutation hill climbing," *ICPR 2004. Proceedings of the 17th International Conference on. Vol. 2. IEEE*.
- [15] R. Fiebrink and I. Fujinaga, "Feature selection pitfalls and music classification," *Proc. 7th Int. Conf. Music Inf. Retrieval ISMIR06*, 2006.
- [16] E. Gómez and P. Herrera, *Comparative analysis of music recordings from western and non-western traditions by automatic tonal feature extraction*. Empirical Musicology Review, 2008.

- [17] C. Harte, M. Sandler, and M. Gasser, "Detecting harmonic change in musical audio," in *Proceedings of the 1st ACM workshop on Audio and music computing multimedia*. ACM, 2006, pp. 21–26.
- [18] K. Hevner, "Experimental studies of the elements of expression in music." *The American Journal of Psychology*, 1936.
- [19] X. Hu, "Music and mood: Where theory and reality meet," in *Proc. iConference*, 2010.
- [20] X. Hu, J. S. Downie, and A. F. Ehmann, "Lyric text mining in music mood classification," *Proc. 10th Int. Conf. Music Inf. Retrieval, ISMIR09*, 2009.
- [21] X. Hu, J. S. Downie, C. Laurier, M. Bay, and A. F. Ehmann, "The 2007 mirex audio mood classification task: Lessons learned," *Proc. 9th Int. Conf. Music Inf. Retrieval, ISMIR08*, 2008.
- [22] S. Jun, S. Rho, B.-J. Han, and E. Hwang, "A fuzzy inference-based music emotion recognition system," *Visual Information Engineering*, 2008.
- [23] P. N. Juslin and J. A. Sloboda, *Handbook of music and emotion: Theory, research, applications*. Oxford University Press, 2010.
- [24] O. Lartillot, D. Cereghetti, K. Eliard, W. J. Trost, M.-A. Rappaz, and D. Grandjean, "Estimating tempo and metrical features by tracking the whole metrical hierarchy," *3rd International Conference on MusicI & Emotion, Jyväskylä*, 2013.
- [25] O. Lartillot, T. Eerola, P. Toivainen, and J. Fornari, "Multi-feature modeling of pulse clarity: Design, validation and optimization," *ISMIR 2008, Automatic Music Analysis and Transcription*, 2008.
- [26] O. Lartillot, *Mirtoolbox 1.5 User's Manual*.
- [27] O. Lartillot and P. Toivainen, "A matlab toolbox for musical feature extraction from audio," in *International Conference on Digital Audio Effects*, 2007, pp. 237–244.
- [28] C. Lauriel, O. Lartillot, T. Eerola, and P. Toivainen, "Exploring relationships between audio features and emotion in music," *ESCOM '09*, 2009.
- [29] M. Levy and M. Sandler, "A semantic space for music derived from social tags," *Austrian Computer Society*, vol. 1, p. 12, 2007.
- [30] Y.-C. Lin, Y.-H. Yang, H. H. Chen, I.-B. Liao, and Y.-C. Ho, "Exploiting genre for music emotion classification," in *ICME*, 2009, pp. 618–621.
- [31] M. McVicar, T. Freeman, and T. D. Bie, "Mining the correlation between lyrical and audio features and the emergence of mood," *Proc. 12th Int. Conf. Music Inf. Retrieval, ISMIR11*, 2011.
- [32] I. Nabney, *NETLAB: algorithms for pattern recognition*. Springer, 2002.

- [33] J. Russell, "A circumplex model of affect," *Journal of Personality and Social Psychology*, 1980.
- [34] P. Saari, T. Eerola, and O. Lartillot, "Generalizability and simplicity as criteria in feature selection: Application to mood classification in music," *IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING*, vol. 19, pp. 1802–1812, 2011.
- [35] Y. Song, S. Dixon, and M. Pearce, "Evaluation of musical features for emotion classification," *Proc. 13th Int. Conf. Music Inf. Retrieval, ISMIR12*, 2012.
- [36] Y.-H. Yang and H. H. Chen, *Music Emotion Recognition*. CRC Press, 2011.



ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ
ΒΙΒΛΙΟΘΗΚΗ



004000121001