



ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

**ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ**

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

**Πολυτροπική Ανάλυση Πληροφορίας για την Εκτίμηση
Συναισθήματος σε Μουσική**

Γρηγόριος Σ. Τσαμαδιάς

Επιβλέποντες: **Γεώργιος Σταμούλης, Καθηγητής Πανεπιστημίου Θεσσαλίας**
Ευάγγελος Σπύρου, Επίκουρος Καθηγητής ΤΕΙ Στερεάς Ελλάδος

ΛΑΜΙΑ

ΟΚΤΩΒΡΙΟΣ 2018

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Πολυτροπική Ανάλυση Πληροφορίας για την Εκτίμηση Συναισθήματος σε Μουσική

Γρηγόριος Σ. Τσαμαδιάς

A.M.: 2153050

ΕΠΙΒΛΕΠΟΝΤΕΣ: Γεώργιος Σταμούλης, Καθηγητής Πανεπιστημίου Θεσσαλίας
Ευάγγελος Σπύρου, Επίκουρος Καθηγητής ΤΕΙ Στερεάς Ελλάδος

«Με ατομική μου ευθύνη και γνωρίζοντας τις κυρώσεις ⁽¹⁾, που προβλέπονται από της διατάξεις της παρ. 6 του άρθρου 22 του Ν. 1599/1986, δηλώνω ότι:

1. Δεν παραθέτω κομμάτια βιβλίων ή άρθρων ή εργασιών άλλων αυτολεξεί **χωρίς να τα περικλείω σε εισαγωγικά** και χωρίς να αναφέρω το συγγραφέα, τη χρονολογία, τη σελίδα. Η αυτολεξεί παράθεση χωρίς εισαγωγικά χωρίς αναφορά στην πηγή, είναι λογοκλοπή. Πέραν της αυτολεξεί παράθεσης, λογοκλοπή θεωρείται και η παράφραση εδαφίων από έργα άλλων, συμπεριλαμβανομένων και έργων συμφοιτητών μου, καθώς και η παράθεση στοιχείων που άλλοι συνέλεξαν ή επεξεργάστηκαν, χωρίς αναφορά στην πηγή. Αναφέρω πάντοτε με πληρότητα την πηγή κάτω από τον πίνακα ή σχέδιο, όπως στα παραθέματα.
2. Δέχομαι ότι η αυτολεξεί **παράθεση χωρίς εισαγωγικά**, ακόμα κι αν συνοδεύεται από αναφορά στην πηγή σε κάποιο άλλο σημείο του κειμένου ή στο τέλος του, είναι αντιγραφή. Η αναφορά στην πηγή στο τέλος π.χ. μιας παραγράφου ή μιας σελίδας, δεν δικαιολογεί συρραφή εδαφίων έργου άλλου συγγραφέα, έστω και παραφρασμένων, και παρουσίασή τους ως δική μου εργασία.
3. Δέχομαι ότι υπάρχει επίσης περιορισμός στο μέγεθος και στη συχνότητα των παραθεμάτων που μπορώ να εντάξω στην εργασία μου εντός εισαγωγικών. Κάθε μεγάλο παράθεμα (π.χ. σε πίνακα ή πλαίσιο, κλπ), προϋποθέτει ειδικές ρυθμίσεις, και όταν δημοσιεύεται προϋποθέτει την άδεια του συγγραφέα ή του εκδότη. Το ίδιο και οι πίνακες και τα σχέδια
4. Δέχομαι όλες τις συνέπειες σε περίπτωση λογοκλοπής ή αντιγραφής.

Ημερομηνία: 01/10/2018

Ο Δηλών



(Υπογραφή)

- (1) «Όποιος εν γνώσει του δηλώνει ψευδή γεγονότα ή αρνείται ή αποκρύπτει τα αληθινά με έγγραφη υπεύθυνη δήλωση του άρθρου 8 παρ. 4 Ν. 1599/1986 τιμωρείται με φυλάκιση τουλάχιστον τριών μηνών. Εάν ο υπαίτιος αυτών των πράξεων σκόπευε να προσπορίσει στον εαυτόν του ή σε άλλον περιουσιακό όφελος βλάπτοντας τρίτον ή σκόπευε να βλάψει άλλον, τιμωρείται με κάθειρξη μέχρι 10 ετών.»

ΠΕΡΙΛΗΨΗ

Στόχος της παρούσας πτυχιακής εργασίας είναι η μελέτη και η εφαρμογή τεχνικών μηχανικής μάθησης στην ανάλυση μουσικής. Αρχικά, παρουσιάζονται σε θεωρητικό επίπεδο η κατηγοριοποίηση των συναισθημάτων, η δομή της μουσικής καθώς και η συσχέτισή της με τα υπολογιστικά συστήματα. Επίσης, παρουσιάζεται ο τρόπος με τον οποίο γίνεται επεξεργασία μουσικών μορφοτύπων για την εξαγωγή και την εκπαίδευση πληροφορίας, μέσω μηχανικής μάθησης. Τέλος, παρουσιάζονται τα εργαλεία που χρησιμοποιήθηκαν και η πρακτική εφαρμογή της προτεινόμενης μεθοδολογίας.

Το δείγμα που δοκιμάστηκε για την λειτουργικότητα της εργασίας αποτελούσαν από κομμάτια καλλιτεχνών/συγκροτημάτων που στην πλειοψηφία τους αντιστοιχούν στις πιο βασικές κατηγορίες στον τομέα της μουσικής. Σε ένα μεγάλο βαθμό η επιλογή τους προέκυψε από την δημοτικότητα τους αλλά κυρίως στον εάν πληρούσαν τις προϋποθέσεις για συλλογή των διαφορετικών πληροφοριών όσον αφορά τα δεδομένα. Η εκπαίδευση έγινε με την χρήση χαρακτηριστικών διαφορετικού μορφότυπου. Πραγματοποιήθηκε σύγκριση μεταξύ εξαγόμενων χαρακτηριστικών από κείμενο έναντι εξαγόμενων χαρακτηριστικών από τα ηχητικά σήματα, χρησιμοποιώντας ένα μοντέλο παλινδρόμησης. Για τα χαρακτηριστικά χρησιμοποιήθηκαν μέθοδοι βραχυπρόθεσμης/ενδιάμεσης εξαγωγής και δημιουργίας σώματος κειμένου για τον ήχο και το κείμενο αντίστοιχα.

ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ: Ανάλυση Κειμένου και Ηχητικού Σήματος μέσω Μηχανικής Μάθησης.

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ: Μουσική, Στοιχείο, Σθένος, Ενέργεια, Χορευτικότητα, Χαρακτηριστικά, Παλινδρόμησης, Μηχανική Μάθηση, Μηχανές Διανυσμάτων Υποστήριξης, Συχνότητα του Όρου στο Κείμενο και Άνιση Κατανομή του Όρου στο Κείμενο, Λανθάνουσα Σημασιολογική Δεικτοδότηση, Μέσο Τετραγωνικό Σφάλμα

ABSTRACT

The purpose of this Thesis is the examination and the implementation of Machine Learning methods in audio analysis. At first there is an introduction at the theory of emotions categorization, the structure of music as well as the correlation between music and computers. Also, there is an allusion to the processing of music formats set out to extract and training information, by using Machine Learning. In the end, there is a description of the tools and the mythologies that applied the theory in practical terms.

The data majority that used to test the functionality of this project be comprised of artist 's/bands' recordings which are appertain to the primary divisions of music. Generally, the selection was done in dependence of the recordings popularity, but the main reason of selection was that the recordings are eligible for the collection of data 's information, which appertain to different categories. The training was become with different format of data. In result of this we made a comparison between sound feature extraction and text feature extraction. To end up in these results we used the model of Regression while for the feature extraction the model of Short-term/Middle-term for the sound and the model of corpus for the text.

SUBJECT AREA: Text and Sound Analysis with Machine Learning

KEYWORDS: Music, Lyrics, Valance, Energy, Danceability, Features, Regression, Machine Learning, Support Vector Machine, Term Frequency-Inverse Document Frequency, Latent Semantic Indexing, Mean Square Error

ΕΥΧΑΡΙΣΤΙΕΣ

Για τη διεκπεραίωση της παρούσας Πτυχιακής Εργασίας, θα ήθελα να ευχαριστήσω τους επιβλέποντες, τον εξωτερικό συνεργάτη Θεόδωρο Γιαννακόπουλο, τον αναπληρωτή καθηγητή Ευάγγελο Σπύρου, τον καθηγητή και πρόεδρο του τμήματος Πληροφορικής Γεώργιο Σταμούλη και τέλος το Εθνικό Κέντρο Έρευνας Φυσικών Επιστημών “Δημόκριτος”, για τη συνεργασία και την πολύτιμη συμβολή τους στην ολοκλήρωση της.

ΠΕΡΙΕΧΟΜΕΝΑ

1. ΕΙΣΑΓΩΓΗ.....	15
1.1 Η Συναισθηματική Επίδραση της Μουσικής στον Άνθρωπο	15
1.1.1 Συναισθηματική Και Μη-Συναισθηματική Προσέγγιση Των Συναισθημάτων	16
1.1.1.1 Το μοντέλο ΙΤΡΑ	17
1.1.1.2 Διαφορετικές Αντιλήψεις	18
1.1.2 Επίδραση της Διάθεσης του Ατόμου και Γενικά Συμπεράσματα	19
2. ΣΥΝΑΙΣΘΗΜΑΤΑ	21
2.1 Δόμηση των Συναισθημάτων	21
2.1.1 Διδιάστατο μοντέλο	22
2.1.2 Πολυδιάστατο μοντέλο.....	23
3. ΜΟΥΣΙΚΗ ΚΑΙ ΥΠΟΛΟΓΙΣΤΙΚΟ ΣΥΣΤΗΜΑ	26
3.1 Μουσική Δομή και Απεικόνιση σε Υπολογιστικό Σύστημα	26
3.2 Μουσικά Μορφότυπα	26
3.2.1 Φυσικό Μορφότυπο	27
3.2.2 Μορφότυπα Παρτιτούρας	28
3.2.3 MIDI Μορφότυπο	31
4. ΔΙΑΧΕΙΡΙΣΗ ΜΟΥΣΙΚΗΣ ΠΛΗΡΟΦΟΡΙΑΣ	33
4.1 Ανάκτησης Μουσικής Πληροφορίας.....	33
4.2 Εξαγωγή Χαρακτηριστικών	34
4.2.1 Εξαγωγή Short-Term Και Mid-Term Χαρακτηριστικών.....	35
4.2.2 Εξαγωγή Τονικών Χαρακτηριστικών	37
4.2.3 Εξαγωγή Χρωματικών και CENS Χαρακτηριστικών.....	39
4.2.4 Εξαγωγή Χαρακτηριστικών από Κείμενο	40
5. ΠΑΛΙΝΔΡΟΜΗΣΗ	42
5.1 Το Στατιστικό Μοντέλο της Παλινδρόμησης	42
5.2 Ορισμοί του Μοντέλου Παλινδρόμησης	42

5.3	Παλινδρόμηση με Μηχανές Διανυσμάτων Υποστήριξης	43
5.3.1	Μαθηματική Προσέγγιση	44
6.	ΜΕΘΟΔΟΛΟΓΙΑ ΚΑΙ ΕΡΓΑΛΕΙΑ	46
6.1	Εργαλεία που Χρησιμοποιήθηκαν για την Υλοποίηση της Εργασίας	46
6.1.1	youtube-dl.....	46
6.1.2	PyLyrics	47
6.1.3	Spotipy	47
6.1.4	Pandas.....	47
6.1.5	pyAudioAnalysis	48
6.1.6	Gensim	52
6.2	Ανάλυση Μεθοδολογίας.....	54
6.2.1	Λήψη και Συλλογή Δειγμάτων Μουσικών Δεδομένων.....	54
6.2.2	Εξαγωγή Χαρακτηριστικών και Εκπαίδευση.....	56
6.2.3	Συλλογή και Σύγκριση Αποτελεσμάτων.....	58
6.3	Αποτελέσματα	58
	ΠΙΝΑΚΑΣ ΟΡΟΛΟΓΙΑΣ	65
	ΣΥΝΤΜΗΣΕΙΣ – ΑΡΚΤΙΚΟΛΕΞΑ – ΑΚΡΩΝΥΜΙΑ	69
	ΑΝΑΦΟΡΕΣ.....	70

ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ

Σχήμα 1.Το μοντέλο ITPRA της θεωρίας των προσδοκιών του Huron στην ροή του χρόνου.....	18
Σχήμα 2.Ο οριζόντιος άξονας αναπαριστά το σθένος και ο κατακόρυφος τη διέγερση..	23
Σχήμα 3.Ο τροχός των συναισθημάτων του Robert Plutchik.....	25
Σχήμα 4.Διαδικασία εξαγωγής Mid-term χαρακτηριστικών.....	36
Σχήμα 5.Τα τρία διαγράμματα περιγράφουν τις φάσεις μείωσης των δειγμάτων των σημάτων x_0 , x_1 και x_2	39
Σχήμα 6.Το επάνω γράφημα είναι αναπαράσταση ενός ηχητικού σήματος όπου ο άξονας των x αντιπροσωπεύει τον χρόνο ενώ ο άξονας των y τη συχνότητα. Το κάτω γράφημα το οποίο ονομάζετε “χρωμόγραμμα” (chromagram) και είναι η χρωματική αναπαράσταση του αρχικού σήματος όπου επίσης ο άξονας των x αντιπροσωπεύει τον χρόνο ενώ ο άξονας των y αντιπροσωπεύει το χρώμα (chroma).	40
Σχήμα 9.Αναπαράσταση του SVM μοντέλου σε υψηλό επίπεδο.....	43
Σχήμα 10.Γραφική αναπαράσταση SVR.....	44
Σχήμα 11.Στο παραπάνω διάγραμμα απεικονίζετε ο κώδικας που εκτελείτε για την λήψη των αρχείων του ήχου καθώς και η λήψη και αποθήκευση σε τοπικό αρχείο του ηλεκτρονικού υπολογιστή.	56

ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ

Εικόνα 1.Απόσπασμα από παρτιτούρα της σύνθεσης “Still Dre” του μουσικού παραγωγού και ράπερ Dr. Dre.	29
Εικόνα 2.Κώδικας XML που περιγράφει την νότα Mi ύφεση (E-flat) η οποία αρχίζει με μέτρο 3 στο φωνητικό κομμάτι.	30
Εικόνα 3.Τμήμα από πλήκτρα πιάνου με ετικέτες με το όνομα του τόνου και τον αριθμό της νότας MIDI.	32
Εικόνα 4.Αρχεία με το μορφότυπο tsv όπου βρίσκονται αποθηκευμένες οι πληροφορίες αρχειοθετημένες σε πίνακες μονοδιάστατους και διδιάστατους αντίστοιχα.	48
Εικόνα 5.Στιγμιότυπου της παράταξης των στοιχείων ενός κομματιού (Τίτλος/Καλλιτέχνης/Όνομα Κομματιού/..) σε μορφή .tsv.	55

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

Πίνακας 1.Κάθε Βραχυπρόθεσμο πλαίσιο αντιπροσωπεύετε από ένα διάνυσμα 34 χαρακτηριστικών όπως παρουσιάζονται στον παρακάτω πίνακα.	49
Πίνακας 2.Ο πίνακας με της παραμέτρους και το αποτέλεσμα του Μέσου Τετραγωνικού Σφάλματος από την εκπαίδευση του ηχητικού σήματος για το Σθένος.	58
Πίνακας 3. Ο πίνακας με της παραμέτρους και το αποτέλεσμα του Μέσου Τετραγωνικού Σφάλματος από την εκπαίδευση του ηχητικού σήματος για την Χορευτικότητα.....	59
Πίνακας 4. Ο πίνακας με της παραμέτρους και το αποτέλεσμα του Μέσου Τετραγωνικού Σφάλματος από την εκπαίδευση του ηχητικού σήματος για την Ενέργεια.	60
Πίνακας 5.Ο πίνακας με της παραμέτρους και το αποτέλεσμα του Μέσου Τετραγωνικού Σφάλματος από την εκπαίδευση των στοίχων για το Σθένος.....	61
Πίνακας 6.Ο πίνακας με της παραμέτρους και το αποτέλεσμα του Μέσου Τετραγωνικού Σφάλματος από την εκπαίδευση των στοίχων για την Χορευτικότητα.	62
Πίνακας 7.Ο πίνακας με της παραμέτρους και το αποτέλεσμα του Μέσου Τετραγωνικού Σφάλματος από την εκπαίδευση των στοίχων για την Ενέργεια.	62

1. ΕΙΣΑΓΩΓΗ

1.1 Η Συναισθηματική Επίδραση της Μουσικής στον Άνθρωπο

Η αναγνώριση συναισθήματος αναμένεται να γνωρίσει μεγάλη ανάπτυξη στις επόμενες γενιές συστημάτων αλληλεπίδρασης ανθρώπου υπολογιστή, καθώς και άλλων πολυμεσικών συστημάτων. Η μελέτη της μουσικής και του συναισθήματος προαπαιτεί την κατανόηση της ανθρώπινης ψυχολογίας καθώς και την επιρροή που έχει η μουσική σε αυτή. Στον τομέα της ψυχολογίας τα συναισθήματα που προκύπτουν από αισθητήριες αντιδράσεις είναι λιγότερο κατανοητά και αποκαλούνται “ψευδο-συναισθήματα” (pseudo-emotions) ή “καλαίσθητα συναισθήματα” (aesthetic emotions). Η ονομασία αυτή προέκυψε διότι επικρατεί η αντίληψη ότι η έκκριση συναισθημάτων μέσα από την τέχνη απέχει σημαντικά από την φυσιολογική. Σύμφωνα με σύγκριση διαφόρων μοντέλων που αναπτύχθηκαν τα συναισθήματα περιγράφονται με όρους ποικίλων διαστάσεων, παρόλα αυτά οι προσεγγίσεις δεν είναι απολύτως συμβατές.

Η μουσική θεωρείται ένα μη φραστικό μέσο επικοινωνίας που προκαλεί ποικίλα είδη συναισθημάτων, τα οποία διαφέρουν από μεταξύ τους, συνεπώς και η συναισθηματική ανταπόκριση διαφέρει για το κάθε είδος μουσικής. Οι διαφορετικές εμπειρίες του ακρατή είναι άλλος ένας παράγοντας που συμβάλλει στην διαφορά των αντιδράσεων των ακουσμάτων και των ερεθισμάτων που του προκαλούν. Από βιολογικής άποψης οι αισθητήριες αντιδράσεις μεταφράζονται σε σήμα ενεργοποίησης διαφόρων τμημάτων του ανθρώπινου σώματος. Για παράδειγμα, ένας έντονος μουσικός τόνος προκαλεί σύσπαση των μυών ενώ αντίθετα ένας χαλαρός μουσικός τόνος χαλαρώνει τους μύες. Από την πλευρά της ανάλυσης των τραγουδιών παρατηρείται ότι η προφορική εκδήλωση συναισθημάτων εκφράζεται πλησιέστερα από τη μουσική. Η μουσική αποτελεί θεραπεία για τον άνθρωπο και μέσω της χρήσης μουσικών οργάνων και φωνητικών ως εργαλεία αναπαράγονται διάφορα συναισθήματα. Ο κάθε καλλιτέχνης εκτελεί το κομμάτι ακολουθώντας τον ρυθμό της μουσικής. Το φυσικό αποτέλεσμα θα ήταν τελείως διαφορετικό εάν ο καλλιτέχνης απλά διάβαζε τους στίχους. Συνεπώς μπορούμε να πούμε ότι η μουσική μεμονωμένα είναι συναισθηματικά επαρκής, ωστόσο μέσω των φωνητικών εκφράζονται πληρέστερα συναισθήματα.

Στην μουσική εκτός από τον ήχο υπάρχουν και άλλοι παράγοντες με μικρότερο βαθμό επιρροής που επιδρούν στο τελικό αποτέλεσμα όπως ο ρυθμός (tempo) και η ηχηρότητα (sonority). Ο ρυθμός για παράδειγμα μπορεί να είναι αργός, γρήγορος, “βαρύς”. Ο συνδυασμός ρυθμών μπορεί να παράγει διαφορετικά είδη συναισθημάτων.

Αντίστοιχα και από την πλευρά των στίχων υπάρχουν επιμέρους παράγοντες που επηρεάζουν το αποτέλεσμα όπως ονοματοποιία, ήχοι φωνηέντων και φυσικά η σημασία των λέξεων. Σημαντικό είναι να αναφερθεί ότι ο σύνθετης δεν αντιλαμβάνεται την μουσική μέσα από την επιστήμη της ψυχολογίας αντιθέτως δημιουργεί βάσει της διάθεσης και των συναισθημάτων του. Μέσα από όλα αυτά ο κλάδος της Ανάκτησης Μουσικής Πληροφορίας απομονώνει την απαραίτητη πληροφορία και την διαθέτει σε διάφορες εφαρμογές προς μελέτη και επεξεργασία. Παρόλο που η μουσική μπορεί να περιγράψει με μαθηματικές σχέσεις, τα μαθηματικά δεν παράγουν μουσική. Ωστόσο πρέπει να επισημάνουμε ότι για να επιτύχουμε ένα αντιπροσωπευτικό αποτέλεσμα στην μελέτη της μουσικής δεν πρέπει να υπέρ-εστιάζουμε την οπτική γωνία του σύνθετη ή αντίστοιχα του ακροατή.

1.1.1 Συναισθηματική Και Μη-Συναισθηματική Προσέγγιση Των Συναισθημάτων

Γενικά επικρατούν δυο αντιλήψεις όσον αφορά την αναπαραγωγή συναισθημάτων μέσα από την μουσική: α) αυτοί που πιστεύουν ότι η μουσική προκαλεί συναισθήματα στον ακροατή (Emotivist) και β) αυτοί που πιστεύουν ότι η μουσική δημιουργεί την ψευδαίσθηση του συναισθήματος (Cognitivist). Όσον αφορά αυτούς που ανήκουν στη δεύτερη κατηγορία, πιστεύουν ότι η έκφραση της μουσικής δεν προσδιορίζεται μέσα από τους ακροατές και ότι αισθάνονται θετικά όταν τους αρέσει ένα κομμάτι ενώ αρνητικά όταν δεν τους αρέσει. Με αλλά λόγια οι ακροατές δεν αναφέρουν ένα κομμάτι ως “χαρούμενο” ή “λυπημένο” επειδή τους προκαλεί το ανάλογο συναίσθημα, αλλά επειδή εκφράζει τέτοιου είδους συναισθήματα. Από την άλλη πλευρά οι υποστηρικτές της πρώτης κατηγορίας πιστεύουν ότι το συναίσθημα ενός τραγουδιού εκφράζεται μέσω του ακροατή.

Έχουν γίνει προσπάθειες για την έκφραση της συναισθηματικής αντίδρασης στη μουσική μέσα από όρους γνωστικής αξιολόγησης. Προκειμένου να αντιμετωπίσουν αυτούς που δυσπιστούν υποστηρίζοντας ότι τα συναισθήματα δεν συμπεριλαμβάνουν αξιολόγηση και ότι άλλοι μηχανισμοί προκαλούν αύξηση στα συναισθήματα που προκύπτουν από την μουσική. Για παράδειγμα μια συναισθηματική αντίδραση στην μουσική μπορεί να είναι μια αντίδραση συμπάθειας. Πιθανολογείται ότι ένα κομμάτι της μουσικής επηρεάζει την συναισθηματική κατάσταση του ακροατή ακόμα και αν αυτός δεν εστιάζει σε αυτή, όπως ακριβώς θα συνέβαινε στην περίπτωση που ένα άτομο ήταν περιτριγυρισμένο από καταθλιπτικά άτομα, η συναισθηματική του κατάσταση θα επηρεαζόταν από αυτή του “όχλου”. Άλλοι ανθρώπινοι μηχανισμοί που δημιουργούν

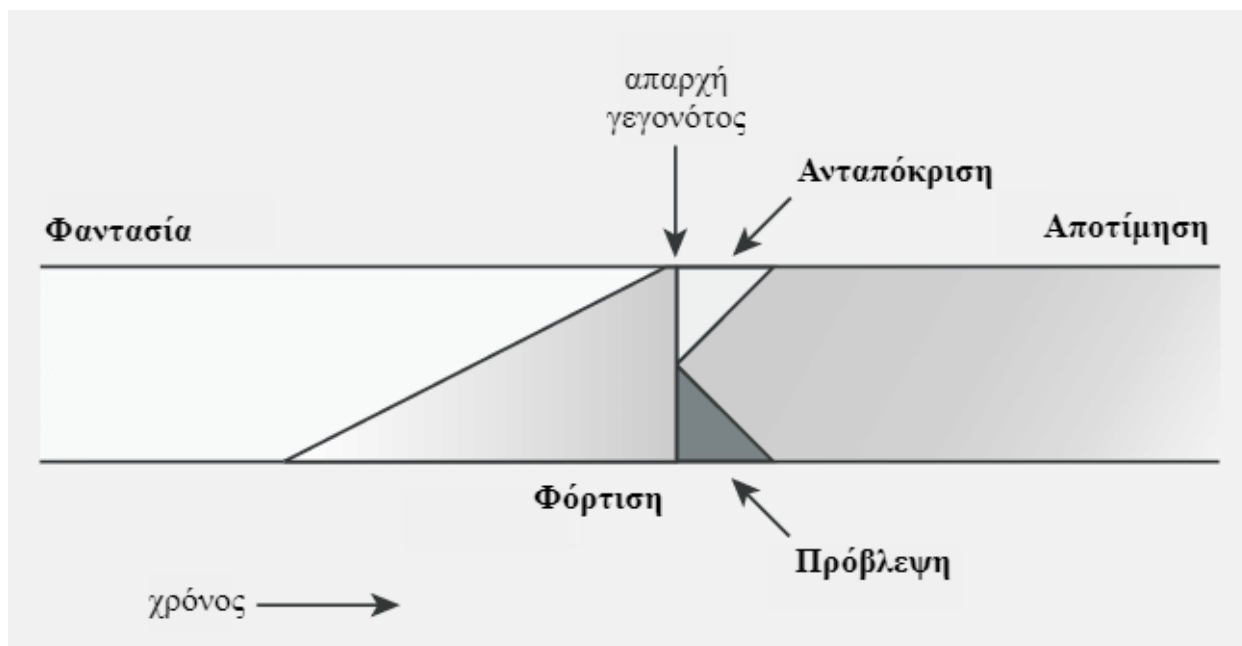
συναισθήματα είναι το τμήμα του εγκεφάλου που είναι υπεύθυνο για τα αντανακλαστικά. Για παράδειγμα, μια παραφωνία μας προκαλεί δυσαρέσκεια ή μεταδοτικότητα (όπως προαναφέρθηκε στο παράδειγμα του όχλου). Η κατάσταση στην οποία βρίσκεται ένα άτομο μπορεί να επηρεάσει με αρνητικά ή θετικά συναισθήματα, η μετατροπή της ακουστικής αφήγησης σε εικονική, αυτοβιογραφική μνήμη που καταγράφει γεγονότα τα οποία χαρακτηρίζονται από συναισθηματική φόρτιση, αλλά και προσδοκίες που έχουν εκπληρωθεί ή έχουν απορριφτεί.

1.1.1.1 Το μοντέλο ITPRA

Μέσα από την θεωρία των προσδοκιών αναπτύχθηκε το μοντέλο ITPRA (Φαντασία/Imagination-Ένταση/Tension-Πρόβλεψη/Prediction-Ανταπόκριση/Reaction-Αποτίμηση/Appraisal) σύμφωνα με τον καθηγητή David Huron όπου αναλύει στο βιβλίο του “Sweet Anticipation: Music and the Psychology of Expectation” το 2006. Συγκεκριμένα αναφέρεται ότι σκοπός της αντίδρασης στη φαντασία είναι η δημιουργία του κινήτρου, που ωθεί τον οργανισμό να συμπεριφέρεται με τρόπο που να αυξάνει την πιθανότητα για μελλοντικά ωφέλιμα αποτελέσματα. Αντίστοιχα η αντίδραση της έντασης, έχει σκοπό την προετοιμασία του οργανισμού για ένα επερχόμενο γεγονός, προσαρμόζοντας τη διέγερση και δίνοντας προσοχή στην αντιστοίχιση του επιπέδου της ανασφάλειας και της σπουδαιότητας του επερχόμενου αποτελέσματος. Με την πρόβλεψη παρέχονται θετικά και αρνητικά κίνητρα ώστε να ενθαρρύνουν τον ακριβή σχηματισμό προσδοκιών. Η ανταπόκριση έχει ως στόχο την αντιμετώπιση της περίπτωσης που εμφανιστεί το χειρότερο σενάριο αναπαράγοντας αντίδραση “άμεσης προστασίας”. Τέλος η αποτίμηση παρέχει θετική και αρνητική ενίσχυση όπου σχετίζεται με την βιολογική τιμή των διαφορετικών τελικών καταστάσεων.

Όπως διακρίνουμε από το σχήμα 1 τα δυο πρώτα στάδια του μοντέλου προκύπτουν πριν την έναρξη του γεγονότος και τα άλλα τρία μετά τον τερματισμό του. Για να ερμηνεύσουμε καλύτερα τις αντιδράσεις στο μουσικό πεδίο ας αναφερθούμε στην ύπαρξη ενός μουσικού γεγονότος (π.χ. δημιουργία μιας σύνθεσης). Η φαντασία αποτελεί ένα προσχέδιο για το μουσικό γεγονός και παροτρύνει τον ακροατή/συνθέτη να συμβάλει. Αντίθετα η συναισθηματική φόρτιση (ή ένταση) σχετίζεται με την ψυχοσωματική κατάσταση του ακροατή/συνθέτη έως ότου λάβει μέρος του μουσικού γεγονότος. Με την πρόβλεψη ο ακρατής οροθετεί τις προσδοκίες του. Με ευαρέσκεια ή δυσαρέσκεια ανταποκρίνεται στο αποτέλεσμα που εμφανίζει το μουσικό αυτό γεγονός. Ενδεχομένως με τη λήξη του μουσικού γεγονότος ο ακροατής να προβάλλει ανάμεικτες

αντιδράσεις για παράδειγμα ένα αποτυχημένο μουσικό γεγονός αλλά με επιτυχημένη πρόβλεψη. Τέλος η αντίδραση που προκύπτει από την αξιολόγηση (ή αποτίμηση) προέρχεται από υψηλού επιπέδου κριτική του επιθυμητού ή μη αποτελέσματος. Συνεπώς κατά την διάρκεια των διαφόρων σταδίων η συναισθηματική κατάσταση του ακροατή/συνθέτη διαφέρει, καθώς το συμβάν μπορεί να ήταν επιτυχημένο αλλά μετά το πέρας του, το άτομο δεν ανταποκρίνεται απαραίτητα με ενθουσιασμό καθώς θα μπορούσε να διακατέχεται από ηρεμία ή κόπωση.



Σχήμα 1. Το μοντέλο ITPRA της θεωρίας των προσδοκιών του Huron στην ροή του χρόνου.

1.1.1.2 Διαφορετικές Αντιλήψεις

Μια άλλη θεωρία υποστηρίζει ότι τα συναισθήματα σχετίζονται με νοητική εκτίμηση. Συγκεκριμένα η συναισθηματική επίδραση της μουσικής στον ακροατή δεν αποτελεί κάποιο συναίσθημα αλλά στην πραγματικότητα είναι ένα κομμάτι ενός συναισθήματος χωρίς να σχετίζεται με την νόηση. Αυτό συγχέει τη διαφορά ανάμεσα σε συναίσθημα και διάθεση, που τυπικά χρησιμοποιείται για την περιγραφή διάχυτων αλλά συγγενικών, μακράς διάρκεις συναισθημάτων χωρίς ξεκάθαρο σκοπό. Κάποιοι διαφωνούν στο ότι η διάθεση προδιαθέτει συγκεκριμένα είδη νόησης, καθώς όταν ζητήθηκε σε πείραμα από τους συμμετέχοντες να ανακαλέσουν μια ανάμνηση θυμού ή λύπης, το επίπεδο διάθεσης-σκέψεων καθόρισε το βαθμό που αντιστοιχίζεται η βιωματική διάθεση. Οι μετρήσεις καταγράφηκαν μετά την εκτέλεση μουσικών κομματιών με περιεχόμενο λύπης, χαράς και θυμού, ωστόσο προέκυψε συσχέτιση μεταξύ διάθεσης και σκέψης

καθώς η μουσική αποσπά την διάθεση, επηρεάζοντας με αυτόν τον τρόπο τις σκέψεις των συμμετεχόντων. Επιπλέον η μουσική με περιεχόμενο λύπης προκαλεί συναισθήματα λύπης λόγω της επίκλησης δυσαρεστών σκέψεων, κάτι που οδηγεί στο συμπέρασμα ότι τα “ρηχά” συναισθήματα εκφράζουν κατάσταση ψυχοσωματικής πάθησης. Ανεξάρτητα με τον εάν τα συναισθήματα που προέρχονται από τη μουσική επάγονται στην διάθεση του ατόμου υπάρχουν αρκετά στοιχεία κατά των φανατικών οπαδών της μη συναισθηματικής προσέγγισης. Οι ψυχολογικές αποκρίσεις όσο και οι νεολογικές σε μουσικό επίπεδο συμφωνούν σε μεγαλύτερο βαθμό με την συναισθηματική προσέγγιση.

Μια άλλη οπτική γωνία σύμφωνα με τους Zentner, M., Grandjean, D. και Scherer, K. το 2008, διαφωνεί στο ότι η συναισθηματική αντίδραση στην μουσική είναι σύνηθες φαινόμενο αλλά αυτό έρχεται σε σύγκρουση με την φυσιολογική σύλληψη των συναισθημάτων. Χαρακτηριστικό παράδειγμα τα συναισθήματα υπέρβασης είναι σύνηθες συνέπεια της μουσικής ακρόασης, ωστόσο η υπέρβαση δεν χαρτογραφείται σε 2D χώρο, ώστε να ορίζεται από τις τιμές διέγερσης και σθένους και είναι ανόμοια από τα πρότυπα συναισθήματα που χρησιμοποιούνται από τις προσεγγίσεις κατηγοριοποιήσεις. Αυτό έχει ως αποτέλεσμα να δημιουργηθεί μια διαφορετική κλάση καλαίσθητων συναισθημάτων που αποκαλούνται καθημερινά ή ωφέλιμα συναισθήματα. Για τα καλαίσθητα συναισθήματα το κομμάτι των συναισθημάτων είναι ξεκάθαρο αλλά τη συμπεριφορά και τα ψυχολογικά δομικά στοιχεία είναι δυσδιάκριτα. Σε έρευνα με αναδρομικές ατομικές αναφορές, οι συμμετέχοντες απαρίθμησαν πόσο συχνά αντιλαμβάνονται και αισθάνονται μεγάλο αριθμό συναισθηματικών εκφράσεων κατά την απόκριση τους στην μουσική. Τα αποτελέσματα των αναλύσεων αποκάλυψαν ότι οι συνήθεις συναισθηματικές αποκρίσεις στη μουσική μπορούν να ομαδοποιηθούν σε μια από τις εννέα κατηγορίες: Κατάπληξη (Wonder), Υπέρβαση (Transcendence), Στοργή (Tenderness), Νοσταλγία (Nostalgia), Γαλήνη (Peacefulness), Ενέργεια (Power), Περίχαρη Δραστηριοποίηση (Joyful Activation), Φόρτιση (Tension), και Λύπη (Sadness). Παρόλα αυτά υπάρχει μια αλληλοεπικάλυψη μεταξύ αυτών των ορισμάτων, τα όποια χρησιμοποιούνται και σε μοντέλα διαστάσεων και κατηγοριών, καθώς επίσης αυτά τα καλαίσθητα συναισθήματα διαφέρουν από τα καθημερινά.

1.1.2 Επίδραση της Διάθεσης του Ατόμου και Γενικά Συμπεράσματα

Ενδέχεται οι ισχυρότερες αποδιώξεις της συναισθηματικής απόκρισης στην μουσική να προέρχονται από μελέτες που αναφέρουν ότι η μουσική επάγεται σε διάθεση. Αναφορές

πιστοποιούν ότι η διάθεση των ακροατών επηρεάζεται υποκειμενικά από την ακρόαση μουσικών κομματιών, εφόσον είναι μετρήσιμα και επιδρούν στην νόηση, παρέχονται αμερόληπτα αποτελέσματα. Αυτό εξακριβώνει η Επίδραση Μότσαρτ, η οποία αναφέρει ότι ακούγοντας συνθέσεις του Μότσαρτ βελτιώνεται η απόδοση σε ελέγχους χωροταξικών ικανοτήτων. Αντίστοιχη βελτίωση μπορεί να αποτελέσει η έκθεση σε πρόσχαρα ερεθίσματα και σε ελέγχους που βαθμολογούν μη χωροταξικές ικανότητες. Σημαντικό έναυσμα αποτελεί η σχέση μεταξύ έκθεσης στην μουσική και νοητικής απόδοσης που επηρεάζεται από την διάθεση του ακροατή και το επίπεδο διέγερσης.

Εφόσον έχουμε λάβει υπόψη όλα τα διαθέσιμα στοιχεία είναι ολοφάνερο ότι η αντίδραση από μουσικά ερεθίσματα οδηγεί σε συναισθήματα πιο σύνθετα από μια απλή έκφραση αρέσκειας ή δυσαρέσκειας. Ωστόσο αυτή η συναισθηματική αντίδραση μπορεί να διαφέρει από το συνήθη ορισμό των συναισθημάτων. Δύσκολη παραμένει η ερμηνεία του γιατί τα άτομα επιλέγουν την ακρόαση μουσικών κομματιών με θλιβερό περιεχόμενο, καθώς οδηγούν σε συναισθήματα θλίψης και λογικά πλαίσια αναμενόμενη θα ήταν η αποφυγή τους. Το βασικότερο ερώτημα είναι αν πρέπει να κατατάσσεται η αντίδραση από μουσικά ακούσματα ως γνήσιο συναίσθημα ή ως ψευδός-συναίσθημα. Από τις έρευνες που πραγματοποιήθηκαν οι συμμετέχοντες δήλωσαν ότι όντως αισθάνθηκαν αυτά τα συναισθήματα οπότε τα αποδεχόμαστε ως “γνησία”.

2. Συναισθήματα

2.1 Δόμηση των Συναισθημάτων

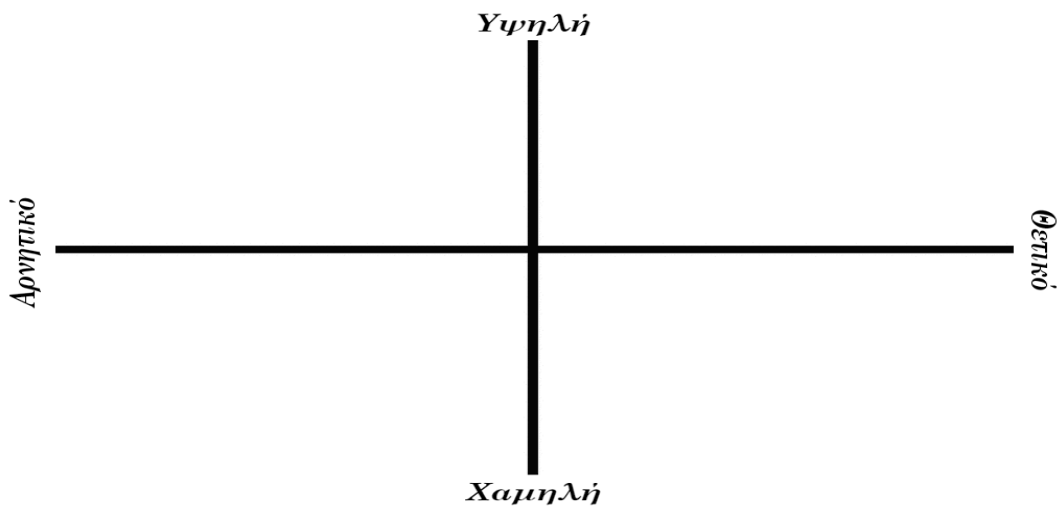
Παρόλο που τα συναισθήματα διακρίνονται σε διάφορες κατηγορίες, ορισμένοι υποστηρίζουν ότι μπορούν να αφομοιωθούν με επιφανειακή περιγραφή, για παράδειγμα συναισθήματα “αρέσκειας” και “δυσαρέσκειας”. Άλλοι αναφέρουν ότι συναισθηματική εμπειρία εκφράζεται με τη χρήση δυο συνεχών διαστάσεων, γνωστό ως μοντέλο circumplex. Όπου οι διαστάσεις αυτές είναι η διέγερση (arousal) με την οποία προσδιορίζεται ο βαθμός κινητοποίησης και περιγράφεται με τις τιμές “υψηλό” και “χαμηλό” καθώς και το “σθένος” (valance) όπου αντίστοιχα καθορίζει τις τιμές δυσαρέσκειας και αρέσκειας και περιγράφεται με τις τιμές “θετικό” και “αρνητικό”. Ποιο συγκεκριμένα για την απεικόνιση του συναισθήματος της ευτυχίας έχει θετικό σθένος και αρκετά υψηλή διέγερση. Αντίθετα συναισθήματα όπως ο φόβος και ο θυμός έχουν αρνητικές τιμές σθένους και υψηλό βαθμό διέγερση. Η διπολική διάσταση του σθένους ερμηνεύει την συνθηματική ανταπόκριση σε αντιδράσεις, όπως το ανοιγοκλείμα των ματιών ως αντανakλαστικό από έναν απρόσμενο κρότο. Ο συνδυασμός υψηλής διέγερσης και θετικού σθένους οδηγεί σε περιορισμένη αντίδραση ωστόσο αν το σθένος ήταν αρνητικό το αποτέλεσμα που προκύπτει είναι μια υπερβολική αντίδραση. Αναμενόμενο ήταν να υπάρχουν και αντίθετες αντιλήψεις που υποστηρίζουν ότι το μοντέλο της διδιάστατης κατηγοριοποίησης δεν είναι επαρκές για την ομαδοποίηση των συναισθημάτων.

Στην έρευνα της μουσικής το δισδιάστατο μοντέλο έχει καθιερωθεί ευρέως με μεγάλη επιτυχία. Σε μια έρευνα χρησιμοποιήθηκε πολυδιάστατο μοντέλο για την μελέτη της δομής που υπόκειται στην συναισθηματική ανταπόκριση σε μουσικά ακούσματα. Ζητήθηκε από τα άτομα που έλαβαν μέρος να ομαδοποιήσουν κομμάτια βασισμένα στην ομοιότητα τους, προσδίδοντας τους συναισθηματική ερμηνεία. Το αποτέλεσμα της ομαδοποίησης έδειξε ότι κυριάρχησαν οι δυο διαστάσεις όπου της ακολουθούσε μια τρίτη που σχετιζόταν με την κινητικότητα. Πιο πρόσφατες έρευνες επιβεβαίωσαν την ύπαρξη τριών διαστάσεων ως ένταση-ομαλότητα (intensity-softness), ικανοποίηση-δυσαρέσκεια (pleasantness-unpleasantness) και επισημότητα-ανεπισημότητα (solemnity-triviality). Οι πρώτες δυο διαστάσεις αντιστοιχούν με το διδιάστατο μοντέλο. Έτσι καταλήγουμε στο συμπέρασμα ότι στα σημεία όπου το διδιάστατο μοντέλο δεν είναι επαρκές, αντικαθίσταται από μοντέλα περισσότερων διαστάσεων.

2.1.1 Διδιάστατο μοντέλο

Αντιλαμβανόμαστε τον προσωπικό συναισθηματικό χώρο ως ένα προσωπικό σημασιολογικό περιβάλλον, όπου συγκεκριμένα συναισθηματικά βιώματα ενός ατόμου λαμβάνουν χώρα. Επιπρόσθετα υποθέτουμε ότι αυτός ο χώρος δεν είναι ίδιος για κάθε άτομο και το δοθέν μέγεθος είναι το μέγιστο ακραίο και ελάχιστο όριο για κάθε πιθανή διάσταση. Για παράδειγμα, στη διάσταση της διέγερσης, αντιστοιχεί το συναίσθημα του θυμού σε ποικίλους βαθμούς. Πιο συγκεκριμένα αυτές οι τιμές αντιστοιχούν στην περίπτωση που έχουμε βιώσει τη μεγαλύτερη ένταση θυμού μέχρι το σημείο που δεν αισθανθήκαμε κανένα απολύτως συναίσθημα. Σε αυτό το σημείο διακρίνουμε ότι αυτές οι τιμές διαφέρουν από άτομο σε άτομο.

Η προσέγγιση μέσω διδιάστατων μοντέλων έχει χρησιμοποιηθεί εκτεταμένα στο πλαίσιο της μελέτης μουσικών και μη ερεθισμάτων. Έπειτα από έρευνες στον τομέα των συναισθημάτων, έχει διαπιστωθεί ότι αυτά τα μοντέλα είναι ακατάλληλα για την ερμηνεία ορισμένων συνήθων συναισθηματικών αντιδράσεων, όπως ασαφή ή ανάμεικτα συναισθήματα. Τα διστακτικά μοντέλα συμπεραίνουν ότι θετικό και αρνητικό σθένος καταλήγει σε αντίθετα άκρα μια δίπολης διάστασης. Συνεπώς θετικά και αρνητικά συναισθήματα είναι αλληλοαποκλειόμενα και δεν μπορούν να συνυπάρξουν ταυτόχρονα. Αντίθετα τα μοντέλα εκτίμησης χώρου υποστηρίζουν ότι η συνύπαρξη μεταξύ αρνητικών και θετικών συναισθημάτων μπορεί να επιτευχθεί κάτω από συγκεκριμένες συνθήκες. Σε έρευνα που έγινε ανατέθηκε στους συμμετέχοντες η ενασχόληση με τυχερά παιχνίδια στα όποια είχαν την δυνατότητα να κερδίσουν ή να χάσουν μεγάλα ή μικρά ποσά. Στην περίπτωση που έχαναν ή κέρδιζαν μεγάλα ποσά αναμενόμενη ήταν η καταγραφή έντονων αρνητικών και θετικών συναισθημάτων αντίστοιχα. Σε περίπτωση που κέρδιζαν ένα μικρό ποσό προέκυπταν θετικά συναισθήματα λόγω της νίκης αλλά παράλληλα και αρνητικά διότι δεν κέρδισαν μεγαλύτερο ποσό. Το αντίστροφο συνέβαινε στην περίπτωση που έχαναν ένα μικρό ποσό τα συναισθήματα που προέκυπταν ήταν αρνητικά λόγω της ήττας αλλά παράλληλα θετικά διότι το ποσό δεν ήταν ιδιαίτερα μεγάλο. Τα συναισθήματα της ευτυχίας και της λύπης είναι αντίθετα σύμφωνα με τη διάταξη που τους προσδίδει το σθένος, ενώ σε επίπεδο νευρικών υποστρωμάτων είναι μερικώς διαχωρίσιμα. Στην περίπτωση που υποθέταμε ότι το σθένος είναι σήμα που προέρχεται από διπολική διάσταση, θα προβλεπόταν ότι χαρά και η λύπη ερμηνεύονται από τον βαθμό ενεργοποίησης ενός υποστρώματος ή δυο υποστρωμάτων, που υπάρχει σύνδεση μεταξύ τους.



Σχήμα 2.Ο οριζόντιος άξονας αναπαριστά το σθένος και ο κατακόρυφος τη διέγερση.

2.1.2 Πολυδιάστατο μοντέλο

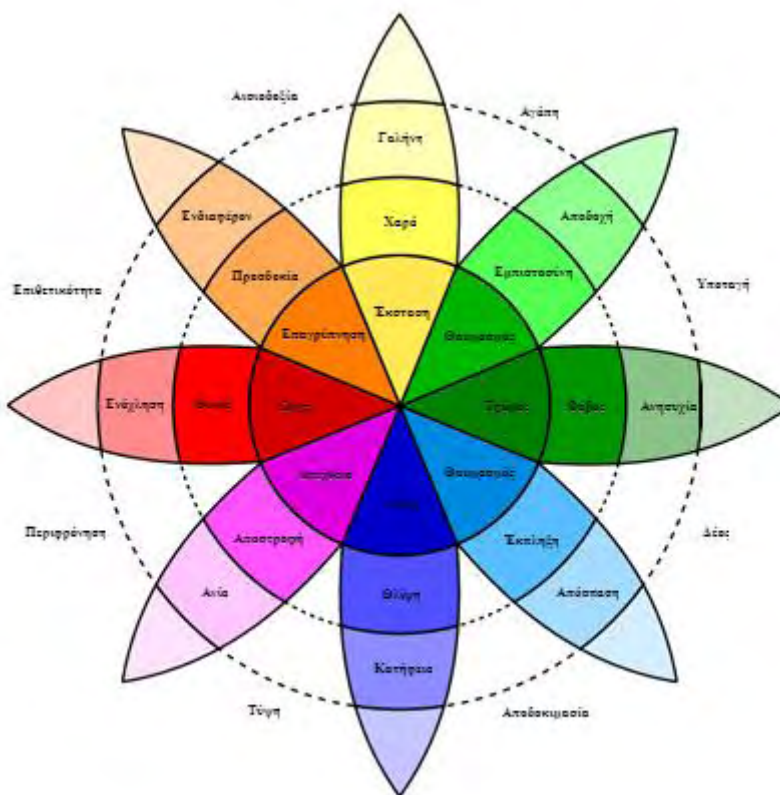
Οι έρευνες δείχνουν ότι τα μουσικά ερεθίσματα αναπαράγουν ανάμεικτα συναισθήματα. Σε πείραμα δυο σταδίων, ζητήθηκε από τους συμμετέχοντες να ακούσουν ηχογραφημένα αποσπάσματα από μουσικές παραστάσεις με περιεχόμενα ευτυχίας, λύπης ή και ανάμιξη των δυο. Στο πρώτο στάδιο η αξιολόγηση των συναισθημάτων έγινε σε δυο διαφορετικές μονοπολικές κλίμακες, ευτυχίας και λύπης, με τιμές από καθόλου έως υπερβολικά. Στο δεύτερο στάδιο η αξιολόγηση έγινε σε διδιάστατο πλέγμα, όπου ο ένας άξονας αντιπροσωπεύει την ευτυχία και ο άλλος τη λύπη. Και στις δυο περιπτώσεις οι συμμετέχοντες κατέγραψαν συγχρόνως χαρά και λύπη όταν τα συναισθήματα ήταν ανάμεικτα. Ένα ακόμα ιδιαίτερο στοιχείο που από τις αναφορές του πειράματος, αναφέρει ότι τα κομμάτια με περιεχόμενο λύπης αποδίδουν σε μεγαλύτερο επίπεδο ανάμεικτα συναισθήματα σε σχέση με αυτά που προσδίδουν χαρούμενο περιεχόμενο. Σε μελέτη όπου χρησιμοποιήθηκαν ελεγχόμενα συναισθήματα από κατεργασμένο υλικό του Μπαχ, μέσα από εφαρμογή ηλεκτρονικού υπολογιστή, σε ήχους χαράς, λύπης και ανάμιξη των δυο, τα αποτελέσματα έδειξαν και εδώ ότι ηχητικά κομμάτια με λύπη στο περιεχόμενο τους εκφράζουν ανάμεικτα συναισθήματα. Μέσα από αυτές τις έρευνες συμπεραίνουμε ότι η μουσική οδηγεί στην δημιουργία ανάμεικτων συναισθημάτων, όπου είναι δύσκολο να διακριθούν μέσα από την εμβέλεια ενός

διπολικού σκέλους αξιολόγησης, με κλίμακα αξιολόγησης από την υπερβολική λύπη στην υπερβολική χαρά.

Εφόσον δεν ήταν επαρκές το μοντέλο των δυο διαστάσεων, για ορισμένες περιπτώσεις αναπτύχθηκαν μοντέλα με τρεις και τέσσερις διαστάσεις, με τη χρήση Καρτεσιανών συντεταγμένων για απεικόνιση των πολυδιάστατων συναισθημάτων. Αρχικά πρέπει να οριστούν οι επιπρόσθετες διαστάσεις. Σύμφωνα με τον Ttnka R. το 2011 μια διάσταση θα μπορούσε να είναι αυτή της προδιάθεσης/σκοπιμότητας (deliberateness ή intentionality), με την οποία συμπληρώνονται ασαφείς τύποι συναισθημάτων όπως η νευρική ανησυχία (anxiety). Μια άλλη διάσταση θα μπορούσε να είναι ο κορεσμός των αναγκών (saturation of needs), καλύπτοντας συναισθήματα όπως η έλλειψη πληρότητας, όπου το άτομο έχει αίσθηση της έλλειψης και από την άλλη πλευρά αίσθηση της ικανοποίησης στην επιτυχία ενός επιθυμητού αποτελέσματος, π.χ. περηφάνεια, θλίψη και νοσταλγία. Μια άλλη διάσταση είναι αυτή των προσδοκιών (expectations), όπου με την προοπτική ως τμήμα προσωπικού βιώματος δεν αποτελεί αποτέλεσμα μόνο γνωστικής διεργασίας, αλλά μια ποιοτική διάσταση που προκύπτει με συναισθηματικές εμπειρίες, περιλαμβάνοντας για παράδειγμα συναισθήματα όπως η ελπίδα και η ανησυχία. Τέλος προτίθεται ως διάσταση η επίγνωση της καταπάτησης ηθικών κανόνων, η οποία εμπεριέχει συναισθήματα όπως η ενοχή. Αυτό το συναίσθημα είναι αποτέλεσμα εμπειριών φόβου και αιφνιδιασμού εφόσον το άτομο αντιληφθεί την παραβίαση που διέπραξε.

Ένα από τα πιο γνωστά πολυδιάστατα μοντέλα είναι ο “τροχός των συναισθημάτων” του σημαντικού ερευνητή στο τομέα της ψυχολογίας Robert Plutchik. Η αναλογία που χρησιμοποίησε ο Plutchik το 1980 μεταξύ ενός πίνακα πιθανών ανθρωπίνων συναισθημάτων και μιας παλέτας χρωμάτων, συσχετίζοντας τα συναισθήματα με τα χρώματα και δηλώνοντας ότι τα συναισθήματα έχουν πολλές “αποχρώσεις” όπου οι διαφορές μεταξύ κάποιων από αυτά είναι αμυδρές, σχεδόν απαρατήρητες. Ο ίδιος σε άλλη εργασία του το 2001 επισημαίνει ότι τα συναισθήματα είναι εσωτερικές προσωπικές καταστάσεις που σπανίως μπορούν να εκφραστούν και να συλληφθούν σε κατανοητή μορφή. Σύμφωνα με το μοντέλο κατηγοριοποίησης του Plutchik υπάρχουν οι εξής πρωτότυπες κατηγορίες: Χαρά (Joy), Εμπιστοσύνη (Trust), Φόβος (Fear) και Έκπληξη (Surprise). Για κάθε ένα από τα τέσσερα αυτά βασικά συναισθήματα υπάρχει και ένας αντίθετος πόλος, Θλίψη (Sadness), Αποστροφή (Disgust), Θυμός (Anger) και Προσδοκία (Anticipation), αντίστοιχα. Οποιοδήποτε άλλο συναίσθημα

μπορεί να εκφραστεί ως μεγαλύτερη ή μικρότερη εκδήλωση των πρωταρχικών αυτών συναισθημάτων ή και ως συνδυασμός τους.



Σχήμα 3.0 τροχός των συναισθημάτων του Robert Plutchik.

Έτσι, καταλήγουμε στο συμπέρασμα ότι με βάση τις ανάγκες που παρουσιάζονται για την επίλυση του ζητούμενου προβλήματος, προσδιορίζεται και ο αριθμός των διαστάσεων του μοντέλου περιγραφής των συναισθημάτων. Ακόμα και το μοντέλο του Plutchik, σε πρακτικό επίπεδο, δεν μπορεί να καλύψει την έκταση της ανθρώπινης συναισθηματικής αντίδρασης. Πολλά μοντέλα χρησιμοποιούν μονάδες που καλύπτουν πλήθος συναισθημάτων, όπως είναι το σθένος. Επιπλέον οι διαστάσεις που θα επιλέξουμε πρέπει να καλύπτουν τις ελλείψεις που δεν αποφεύγονται με τη χρήση άλλων προσεγγιστικών μοντέλων. Η εύρεση κατάλληλων διαστάσεων αποτελεί ένα κομμάτι που βρίσκεται ακόμα σε ερευνητικό στάδιο, καθώς ο ανθρώπινος οργανισμός είναι ένας πολύπλοκος μηχανισμός και δύσβατος σε πολλά σημεία του. Παρόλα αυτά δεν αποτρέπει την έρευνα αλλά αντίθετα δημιουργεί πεδία μελέτης για την ανακάλυψη νέων πληροφοριών.

3. Μουσική και Υπολογιστικό Σύστημα

3.1 Μουσική Δομή και Απεικόνιση σε Υπολογιστικό Σύστημα

Η μουσική προσδιορίζεται ως μεταβολή ηχητικών συχνοτήτων ως προς το χρόνο. Ένα μουσικό κομμάτι αποτελείται από ένα σύνολο ηχητικών δειγμάτων που έχουν υποστεί επεξεργασία και το αποτέλεσμα είναι η μουσική σύνθεση. Ηχητικό δείγμα δεν αποτελούν μόνο ήχοι που παράγονται από τα παραδοσιακά μουσικά όργανα, συμπεριλαμβανομένων απλών καθημερινών ήχων όπως ο ήχος της βροχής, των κυμάτων της θάλασσας κλπ..

Το ηχητικό αποτέλεσμα που λαμβάνει ο άνθρωπος μέσω της ακοής προέρχεται από ταλαντώσεις της ατμοσφαιράς και των αντικείμενων που βρίσκονται στο περιβάλλον γύρω του. Όταν αυτοί οι ήχοι είναι “χαοτικοί” και συγχαιρόμενοι, το αποτέλεσμα τους αποκαλείται “θόρυβος”. Η ευχαρίστηση που λαμβάνουμε από τον θόρυβο είναι αμυδρή. Ωστόσο ορισμένες ηχητικές πηγές και πιο συγκεκριμένα τα μουσικά όργανα αναπαράγουν συμμετρικά ηχητικά κύματα ακολουθώντας μοτίβα. Αυτού του είδους οι ήχοι αναπαράγουν μουσική που ικανοποιούν τις απαιτήσεις των ακροατών. Χαρακτηριστικό είναι το πείραμα της τοποθέτησης άμμου σε μεταλλικά πιάτα, από το οποίο προκύπτει το συμπέρασμα ότι όταν τα πιάτα πάλλονται, από την πίεση που τους ασκήθηκε με ένα δοξάρι από βιολί, η άμμος παρατάθηκε σε γεωμετρική διάταξη. Κατά την διάρκεια των ποικίλων αρμονικών διαταραχών που υπέστησαν τα μεταλλικά πιάτα, αυτή η γεωμετρική διάταξη πλησιάζει σε αυτό που αποκαλούμε μουσική.

3.2 Μουσικά Μορφότυπα

Οι σύγχρονες ψηφιακές μουσικές βιβλιοθήκες αποτελούνται από δεδομένα σε μορφή κειμένου, εικόνας και ήχου. Ανάμεσα σε όλες αυτές τις μορφές πληροφορίας βασισμένες σε πολυμέσα, τα μουσικά δεδομένα παρουσιάζουν ένα ευρύ πλήθος προβλημάτων καθώς η πληροφορία παρουσιάζεται με διάφορα μορφότυπα δεδομένων. Τα διάφορα αυτά μορφότυπα εξαρτώνται από συγκεκριμένες εφαρμογές, οι οποίες διαφέρουν στην μορφή και το περιεχόμενο. Τρία είναι τα πιο δημοφιλή μορφότυπα: η παρτιτούρα, το ακατέργαστο ηχητικό και η Ψηφιακή Διασύνδεση Μουσικών Οργάνων. Για την υλοποίηση της Πτυχιακής εργασίας χρησιμοποιήθηκε μόνο το ακατέργαστο ηχητικό φυσικό μορφότυπο, αλλά παρακάτω περιγράφονται και τα άλλα δύο για να υπάρχει μια σφαιρική κάλυψη επί του θέματος. Το μορφότυπο παρτιτούρας (score format) περιέχει πληροφορίες σχετικά με τις νότες, όπως ο μουσικός χρόνος έναρξης

(musical onset time), η τονικότητα (pitch), η διάρκεια (duration) καθώς περεταίρω στοιχεία που αφορούν τη δυναμική (dynamics) και την ενίσχυση της διάρκειας της έντασης (agogics). Το ακατέργαστο ηχητικό φυσικό μορφότυπο (audio format) κωδικοποιεί την κυματομορφή ενός ηχητικού σήματος όπως έχει αναφερθεί προηγουμένως, η οποία χρησιμοποιείται για την ηχογράφηση σε ψηφιακούς δίσκους. Εξίσου διαδεδομένη είναι η Ψηφιακή Διασύνδεση Μουσικών Οργάνων (Musical Instrument Digital Interface/MIDI format) η οποία θεωρείται ως “υβρίδιο” των δυο προηγούμενων μορφοτύπων περιλαμβάνοντας πληροφορία βάση περιεχομένου όπως έναρξη νότας και τονικότητα αλλά επίσης κωδικοποιεί την ενίσχυση της διάρκειας της έντασης και δυναμική λεπτότητα μιας συγκεκριμένης ερμηνείας.

3.2.1 Φυσικό Μορφότυπο

Από φυσικής όψεως τα μουσικά όργανα παράγουν παλμούς στην ατμόσφαιρα με τη μορφή ηχητικών κυμάτων τα οποία είναι συνήθως περιοδικά, και τα αποκαλούμε “τόνους”. Ο τόνος εκφράζεται από τρεις παραμέτρους την τονικότητα (pitch), την ένταση (intensity) και την χροιά ήχου (tone quality). Η τονικότητα είναι η περιγραφή του τόνου στον άξονα των συχνοτήτων (αριθμός των μουσικών κυμάτων ανά δευτερόλεπτο με μονάδα μέτρησης τα Hertz) ή διαφορετικά ο τρόπος με τον οποίο η ανθρώπινη ακοή αντιλαμβάνεται το μήκος κύματος του παραγόμενου ήχου. Όταν περιγράψουμε τους τόνους μέσω της τονικότητας αποκαλούνται και ως “νότες”. Η χαμηλή τονικότητα έχει μεγάλο μήκος κύματος ενώ η υψηλή έχει μικρό. Η ένταση είναι αυτή που καθορίζει το πόσο δυνατός η απαλός είναι ο ήχος (μονάδα μέτρησης τα decibels). Η τονικότητα αναπαριστά το μήκος κύματος, η ένταση αναπαριστά το πλάτος. Υψηλού πλάτους ηχητικά κύματα αναπαράγουν ήχους ισχυρής έντασης, αντίθετα χαμηλού πλάτους ηχητικά κύματα αναπαράγουν χαμηλής έντασης ήχους. Η τελευταία παράμετρος, η τονική ποιότητα, αποκαλείται και “τονικό χρώμα” ή “ηχόχρωμα”. Στην ανθρώπινη ακοή υπάρχει η δυνατότητα διάκρισης του ήχου όταν μια φλογέρα και ένα βιολί αναπαράγουν την ίδια νότα, αυτή η διαφορά εκφράζεται μέσα από το χρώμα. Ο τόνος που αναπαράγεται από το βιολί είναι πλούσιος και παράλληλα εκφράζει θλίψη, συγκριτικά με το φλάουτο όπου είναι ήρεμος και λιγότερο περίπλοκος. Κάνοντας την συσχέτιση με τις άλλες δυο παραμέτρους η τονική ποιότητα αντιστοιχίζεται με την ίδια την κυματομορφή, συνεπώς αναπαριστάται από τη γραφική της παράσταση μια συνάρτησης.

Το τμήμα του εργαλείου σύνθεσης που είναι υπεύθυνο για την παραγωγή του ήχου ονομάζεται “ταλαντωτής”. Ο ταλαντωτής έχει την δυνατότητα να αναπαράγει την συχνότητα και την αρμονική(επιμέρους ήχοι όπου οι συχνότητες τους αναπαριστούν ακέραιο αριθμό πολλαπλασιασμένο με τη θεμελιώδη συχνότητα). Αυτό σημαίνει ότι ο ταλαντωτής και τη δυνατότητα να αναπαράγει ήχους με ευδιάκριτη τονικότητα, καθώς επίσης έχει την δυνατότητα να μεταδώσει συγκεκριμένο ηχόχρωμα σε αυτές τις νότες. Η δεύτερη ιδιότητα παρέχει την δυνατότητα στον ταλαντωτή να αναπαράγει τις χαρακτηριστικές πολυσύνθετες κυματομορφές σε συσχέτιση με όργανα όπως βιολί πιάνο κλπ. Κατά τη διεργασία της προσθετικής σύνθεσης, η ζητούμενη κυματομορφή σχηματίζεται από την πρόσθεση των αρμονικών κυμάτων με τη θεμελιώδη συχνότητα. Αντίστοιχα λειτουργεί και η αντίστροφη διεργασία της αφαιρετικής σύνθεσης, στην οποία από μια κυματομορφή με επιλεγμένα φίλτρα εξάγονται συγκεκριμένες συχνότητες. Για να είναι επαρκής η σύνθεση ενός ήχου πρέπει να λάβουμε υπόψη την περιβάλλουσα (χαρακτηριστικό που αναπτύσσει ένας ήχος στην πάροδο του χρόνου) αυτού του ήχου. Η περιβάλλουσα αποτελείται από τις εξής φάσεις, α) η αργή ή γρήγορη αύξηση της έντασης ή δράση (Attack), β) η διατήρηση στο μέγιστο επίπεδο έντασης (Hold), γ) η αρχή της μείωσης της έντασης ή μερικής εξασθένησης (Decay), δ) η διατήρηση ενός μείωση επιπέδου έντασης (Sustain) και ε) τέλος την δραστική μύωση και μηδενισμό της έντασης ή αποδέσμευσης (Release). Με την περιβάλλουσα μπορούμε να δημιουργήσουμε ένα σχεδόν ρεαλιστικό αντίγραφο του ήχου των παραδοσιακών οργάνων και όχι μόνο, καθώς με τα διάφορα φίλτρα/ εργαλεία παραμόρφωσης παρέχεται η δυνατότητα δημιουργίας μοναδικών νέων ήχων. Σε ένα υπολογιστικό σύστημα όλα όσα αφορούν τα σήμα του ήχου σε φυσικό επίπεδο, ερμηνεύονται και επεξεργάζονται ως μαθηματικές συναρτήσεις οι οποίες γίνονται εύκολα αντιληπτές βάσει της δομής του.

3.2.2 Μορφότυπα Παρτιτούρας

Η παρτιτούρα δίνει μια συμβολική περιγραφή αυτό που αναφέρεται στην Δυτική κλασική μουσική ως “κομμάτι μουσικής”. Η παρτιτούρα κωδικοποιεί το μουσικό έργο σε επίσημη γλώσσα και απεικονίζεται σε μορφή εικονογράφησης και κειμένου, η οποία επιτρέπει στον μουσικό να δημιουργήσει μια παράσταση ακολουθώντας τις δοθείσες οδηγίες. Στην παρτιτούρα η μουσική αναπαρίσταται από νότες , οι οποίες με τις σειρά τους δίνουν όρους από χαρακτηριστικά όπως η τονικότητα, ο μουσικός χρόνος έναρξης καθώς και άλλες μορφές όπως αναφέρθηκαν και παραπάνω. Ο ρυθμός

συγκεκριμενοποιείται από σύμβολα σε μορφή κειμένου όπως το *Andante con moto* για οδηγίες “καθολικού ρυθμού” (global tempo) και για “τοπικού ρυθμού” (local tempo) επιταχυνόμενος (*accelerando*) ή επιβραδυνόμενος (*ritardando*). Η παρτιτούρα μπορεί να ερμηνευτεί ως ένα μέσω οδηγού που προαπαιτεί γνώσεις και εμπειρία από το μουσικό για τη δημιουργία του επιθυμητού ήχου. Υπάρχει μεγάλη ελευθερία ερμηνείας που οδηγεί σε αλλαγές στο ρυθμό, την άρθρωση ή την δυναμική. Ακόμα και στο επίπεδο της νότας μπορεί να υπάρξουν παραλλαγές στην εκτέλεση με την χρήση τεχνικών όπως το τρίλισμα (*trill*), αρπέτζιο (*arpeggio*), ή την επέριση (*grace note*).



Εικόνα 1. Απόσπασμα από παρτιτούρα της σύνθεσης “Still Dre” του μουσικού παραγωγού και ράπερ Dr. Dre.

Πολλοί κώδικες προτάθηκαν για την αναπαράσταση της παρτιτούρας σε bits ώστε να είναι αναγνώσιμη από έναν υπολογιστή. Ένα τέτοιο μορφότυπο το οποίο είναι και το πιο διαδεδομένο είναι το MusicXML με επέκταση αρχείου .xml, το οποίο αποτελεί μια σύγχρονη επινόηση για καθολική μετάφραση Δυτικής μουσικής σημειογραφίας. Το MusicXML περιγράφει το πως οι νότες, το μετρό, το πεντάγραμμα καθώς και άλλα, παριστάνονται σε μια παρτιτούρα.

```
<note>
  <pitch>
    <step>E</step>
    <alter>-1</alter>
    <octave>5</octave>
  </pitch>
  <duration>24</duration>
  <tie type="start"/>
  <lyric>
    <syllabic>end</syllabic>
    <text>meil</text>
    <extend/>
  </lyric>
</note>
```

Εικόνα 2.Κώδικας XML που περιγράφει την νότα Mi ύφεση (E-flat) η οποία αρχίζει με μέτρο 3 στο φωνητικό κομμάτι.

Ένας τρόπος για την αναπαραγωγή ψηφιακής παρτιτούρας είναι η χειροκίνητη εισαγωγή του μορφότυπου όπως το MusicXML, το οποίο είναι κοπιαστικό και μπορεί να δημιουργήσει σφάλματα. Τα λογισμικά σημειογραφίας, αναφερόμενα και ως *scorewriters*, επιτρέπουν στον χρήστη την συγγραφή και επεξεργασία ψηφιοποιημένης παρτιτούρας, όπου μπορεί να εισάγει και να επεξεργαστεί τις νότες κατάλληλα από το πληκτρολόγιο και το ποντίκι του υπολογιστή ή από κάποιο ηλεκτρονικό όργανο συνήθως αρμόνιο. Πολλά από αυτά τα προγράμματα σημειογραφίας, συμπεριλαμβανομένων και των δυο πιο διάσημων από αυτά Finale και Sibelius, υποστηρίζουν μετατροπέα αρχείων σε μορφότυπο MusicXML.

Ένας άλλος συνήθης τρόπος για την αναπαραγωγή ψηφιακής παρτιτούρας είναι η σάρωση της παρτιτούρας, κάτι που την μετατρέπει σε ένα σύνολο από ψηφιοποιημένες εικόνες. Σε αυτό το στάδιο ο υπολογιστής λαμβάνει υπόψη ότι αυτές οι εικόνες είναι ένα σύνολο από εικονοστοιχεία ή δείγματα χρωμάτων και δεν αντιλαμβάνεται την σημασιολογία της παρτιτούρας. Στο επόμενο στάδιο οι ψηφιακές εικόνες πρέπει να μεταφραστούν περεταίρω σε ένα πρότυπο κωδικοποιημένο σχήμα όπως το MusicXML, το οποίο αντανakλά την σημασιολογία των συμβόλων της παρτιτούρας όπως νότες, παύσεις ή μουσικά κλειδιά. Η διεργασία σάρωσης και της μετατροπής της παρτιτούρας σε μορφότυπο επεξεργάσιμο από ένα υπολογιστή είναι γνωστή ως Οπτική Μουσική Αναγνώριση (Optical Music Recognition/OMR). Παρόλα αυτά τίθενται προβλήματα στην αναγνώριση των συμβόλων, κατά το σκανάρισμα, ακόμα και ένα μικρό λάθος όπως μια τελεία μπορεί να οδηγήσει σε λανθασμένη ερμηνεία της παρτιτούρας.

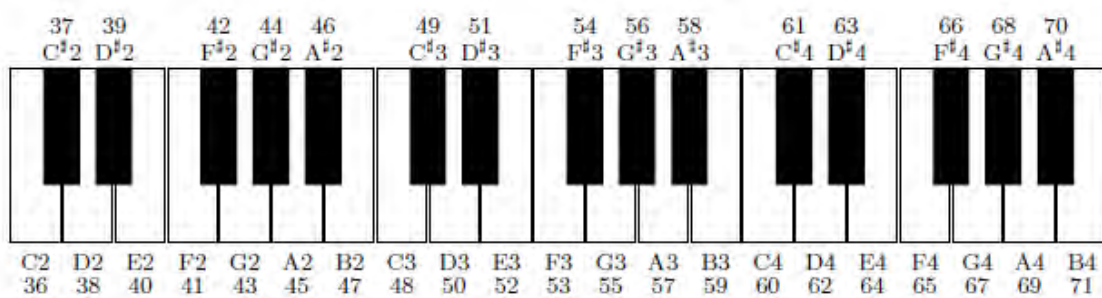
3.2.3 MIDI Μορφότυπο

Όπως προαναφέρθηκε η αναπαράσταση της μουσικής σε MIDI είναι ένας συνδυασμός των δυο παραπάνω μεθόδων παρτιτούρας και ηχητικής αναπαράστασης. Ωστόσο είναι αρκετά περιορισμένη μέθοδος ιδιαίτερα στην αναπαράσταση του ηχοχρώματος του ήχου. Στις μέρες μας το MIDI χρησιμοποιείται ως ένα από τα πιο συνηθισμένα ψηφιακά μουσικά μορφότυπο σε Λογισμικά Μουσικής Σημειογραφίας (Notation Software), Μουσικής Εγγραφής (Sequencer) και Μουσικής Εκπαίδευσης. Αρχικά δημιουργήθηκε ως ένα βιομηχανικό πρότυπο, το οποίο λάμβανε ψηφιακά ηλεκτρονικά μουσικά όργανα, από διαφορετικούς κατασκευαστές, για να παράγει ένα πρότυπο μουσικό αποτέλεσμα μέσα από τον συνδυασμό τους. Μετέπειτα η αρχική προδιαγραφή του MIDI εξελίχθηκε με την εισαγωγή μορφότυπου αρχείων, το Πρότυπο MIDI Αρχείο (Standard MIDI File/SFM), το οποίο περιγράφει πως τα δεδομένα MIDI πρέπει να αποθηκεύονται σε έναν υπολογιστή.

Από πλευράς χρήστη, με το MIDI ο μουσικός έχει τον έλεγχο του μουσικού οργάνου σε πραγματικό χρόνο. Η διαδικασία έλεγχου είναι ίδια με αυτή ενός απλού μουσικού οργάνου χρησιμοποιώντας το πληκτρολόγιο ή ακόμα και κάποιου ηλεκτρονικού οργάνου που είναι συνδεδεμένο με τον υπολογιστή. Αντί να έρχεται σε φυσική επαφή με το μουσικό όργανο ο μουσικός παράγει το ίδιο ηχητικό αποτέλεσμα μεταδίδοντας τα κατάλληλα MIDI μηνύματα, τα όποια περιέχουν κωδικοποιημένα στοιχεία όπως το πότε αρχίζει και πότε τελειώνει η μελωδία που παράγει μια νότα, την ταχύτητα της και άλλες σχετικές πληροφορίες. Από αυτό συμπεράνουμε ότι το MIDI δεν αντιπροσωπεύει τον ήχο απευθείας αλλά εκπροσωπεί την πληροφορία εκτέλεσης, κωδικοποιώντας την εντολή για το πώς ο μουσικός αναπαρήγαγε τον ηχητικό αποτέλεσμα.

Ο Müller M. το 2007 στο βιβλίο του "Information Retrieval for Music and Motion" αναλύει λεπτομερώς τον τρόπο λειτουργίας του MIDI. Συγκεκριμένα αναφέρει ότι το μορφότυπο επιτρέπει στους χρήστες να ανταλλάσουν αρχεία ανεξαρτήτως του λειτουργικού συστήματος του υπολογιστή. Επίσης προσφέρει μια βάση για μια αποδοτική διαδικτυακή κατανομή των μουσικών δεδομένων σε SFM μορφότυπο, συμπεριλαμβάνοντας ένα μεγάλο αριθμό από διαδικτυακές ιστοσελίδες που πωλούν και ανταλλάσουν τέτοιου είδους δεδομένα. Ένα αρχείο MIDI περιέχει τη λίστα από τα μηνύματα μαζί με χρονοσφραγίδες (timestamps), οι οποίες είναι απαραίτητες για τον προσδιορισμό του χρόνου ζωής των μηνυμάτων. Τα μηνύματα που έχουν κυρίαρχο ρόλο είναι οι εντολές εκκίνησης και τερματισμού της νότας (note-on και note-off). Ένα τέτοιο μήνυμα μπορεί να περιέχει στοιχεία όπως τον αριθμό MIDI της νότας, τιμή για το

key velocity, προδιαγραφή καναλιού καθώς και μια χρονοσφραγίδα. Ο αριθμός MIDI της νότας είναι ακέραιος από το 0 έως το 127 και αντιστοιχεί στον κωδικοποιημένο τόνο της νότας. Ένα απλό ακουστικό πιάνο έχει 88 κλειδιά και νότες από A0 ή Λα 0ης οκτάβας έως C8 ή Ντο 8ης οκτάβας, ο κωδικοποιημένος αριθμός νότας MIDI σε αύξουσα σειρά είναι από το C0 ή ντο μηδενικής οκτάβας έως G#9 ή σολ δίεση ενάτης οκτάβας. Ομοίως και ο αριθμός δυναμικής είναι ακέραιος από το 0 έως το 127, όπου ελέγχει την ένταση του ήχου. Στην περίπτωση ενός γεγονότος εκκίνησης νότας προσδιορίζει την ένταση, ενώ σε ένα γεγονός τερματισμού νότας ελέγχει την μείωση του ήχου κατά τη διάρκεια του σταδίου απελευθέρωσης της νότας. Επίσης ο αριθμός δυναμικής εξαρτάται από το όργανο εκτέλεσης. Το MIDI κανάλι είναι ένας ακέραιος από το 0 έως το 15, όπου προτρέπει το συνθεσάιζερ να χρησιμοποιήσει το όργανο που του ανατέθηκε. Τέλος η χρονοσφραγίδα είναι ένας ακέραιος που αντιπροσωπεύει τους παλμούς αναμονής από την εκτέλεση εντολής εκκίνησης νότας.



Εικόνα 3.Τμήμα από πλήκτρα πιάνου με ετικέτες με το όνομα του τόνου και τον αριθμό της νότας MIDI.

4. Διαχείριση Μουσικής Πληροφορίας

4.1 Ανάκτησης Μουσικής Πληροφορίας

Για την μουσική υπάρχει ένα μεγάλο πλήθος ψηφιοποιημένων δεδομένων καθώς και μια ποικιλία σχεσιακών αναπαραστάσεων δεδομένων, τα όποια περιγράφουν την μουσική σε ποικίλα σημασιολογικά επίπεδα. Συνήθως οι συλλογές ψηφιακής μουσικής περιέχουν ένα μεγάλο αριθμό σχετιζόμενων ψηφιακών αρχείων για ένα μουσικό έργο, τα όποια έχουν δοθεί σε διάφορα μορφότυπα και σε πολλαπλές υλοποιήσεις. Για παράδειγμα στην πέμπτη συμφωνία του Μπετόβεν, μια ψηφιακή μουσική βιβλιοθήκη μπορεί να περιέχει σαρωμένες σελίδες από κάποια συγκεκριμένη μουσική θεματολογία. Επίσης θα μπορούσε να έχει δοθεί ως ψηφιακό μορφότυπο αρχείου μουσικής σημειογραφίας, το οποίο κωδικοποιεί την διάταξη της σελίδας μιας παρτιτούρας σε αναγνώσιμη μορφή. Επιπρόσθετα η βιβλιοθήκη θα μπορούσε να περιέχει ποικίλες ηχογραφήσεις ψηφιακών δίσκων από ερμηνείες σύγχρονων καλλιτεχνών του χώρου. Αυτές οι ερμηνείες συχνά αποκλίνουν σε σχέση με τον ρυθμό, τη μουσική άρθρωση, τη μελωδικότητα, ή την ενορχήστρωση.

Σύμφωνα με την απεικόνιση του παραπάνω παραδείγματος, υπάρχουν ποικίλες ψηφιακές ενδείξεις μιας μουσικής σύνθεσης για διαφοροποίηση ως προς τα μορφότυπα και το περιεχόμενο. Στον τομέα της Ανάκτησης Μουσικής Πληροφορία (Music Information Retrieval/MIR), έχουν καταβληθεί προσπάθειες με στόχο την ανάπτυξη τεχνολογιών που επιτρέπουν στο χρήστη να έχει πρόσβαση και να εξετάζει τη μουσική από όλες τις όψεις. Για παράδειγμα, κατά την αναπαραγωγή μιας ηχογράφησης δίσκου, μια ψηφιακή συσκευή αναπαραγωγής μουσικής μελλοντικά αναπαριστά το ανάλογο μουσικό αποτέλεσμα όσο επισημαίνεται η τωρινή θέση αναπαραγωγής εντός του αποτελέσματος. Κατ' απαίτηση, επιπρόσθετες πληροφορίες σχετικά με την μελωδική και αρμονική εξέλιξη ή τον ρυθμό αυτόματος παρουσιάζονται στον χρήστη. Το κατάλληλο περιβάλλον χρήστη επιδεικνύει τη μουσική δομή από το τρέχον μουσικό κομμάτι και επιτρέπει την χρήση άμεσου άλματος σε οποιοδήποτε τμήμα της ηχογράφησης χωρίς την γρήγορη κίνηση του κομματιού προς τα μπροστά ή πίσω. Επιπρόσθετα ο ακροατής είναι εξοπλισμένος με μια μηχανή αναζήτησης που του δίνει τη δυνατότητα να εξερευνήσει συλλογές μουσικών κομματιών με ποικίλους τρόπους. Συγκεκριμένα ο χρήστης δημιουργεί ένα ερώτημα δίνοντας συγκεκριμένες τιμές ενός πλήθους από νότες ή αρμονικές ή ρυθμικές ακολουθίες, σφυρίζοντας την μελωδία ή επιλέγοντας ένα μικρό τμήμα από ένα ηχογραφημένο δίσκο. Αφού το σύστημα επεξεργαστεί τα δεδομένα, παρέχει στο χρήστη μια ταξινομημένη λίστα από τα

διαθέσιμα κομμάτια που σχετίζονται μουσικά με το ερώτημα του. Ανάμεσα στα αποτελέσματα δεν βρίσκεται μόνο το πρωτότυπο κομμάτι, αλλά μπορεί να περιέχονται διαφορετικές ερμηνείες και διασκευές του. Μια προηγμένη μηχανή αναζήτηση έχει την δυνατότητα να αναγνωρίσει το θέμα ακόμα και αν το ερώτημα προέρχεται από παραλλαγή του πρωτότυπου κομματιού.

Παρόλη την πρόοδο που έχει γίνει στην ανάπτυξη των προηγμένων συσκευών αναπαραγωγής μουσικής, υπάρχουν ακόμα άλυτα προβλήματα στην αναζήτηση και ανάκτηση της μουσικής, τα όποια προκύπτουν από την ετερογένεια και την πολυπλοκότητα των μουσικών δεδομένων. Ο όρος “βάσει του περιεχομένου” (content based) σημαίνει ότι στη σύγκριση των μουσικών δεδομένων, το σύστημα χρησιμοποιεί μόνο ακατέργαστα δεδομένα, προκειμένου να στηρίζεται σε δεδομένα που έχουν παραχθεί χειροκίνητα, όπως λέξεις κλειδιά ή άλλες συμβολικές περιγραφές. Αντίστοιχα η ανάκτηση βάσει του κειμένου (text-based) των μουσικών εγγράφων χρησιμοποιεί το όνομα του συνθέτη, τον αριθμό μουσικής σύνθεσης, ή στίχους που μπορούν να διαχωριστούν μέσα από κλασικές τεχνικές ανάκτησης από βάσεων δεδομένων. Φυσικά, η πλήρης ανάκτηση βάσει του περιεχομένου αποτελεί δύσκολο ερευνητικό κομμάτι. Τα βασικά προβλήματα του MIR αφορούν την διαχώριση και ανάκτηση μουσικής πληροφορίας, την αυτόματη αναγνώριση και κατηγοριοποίηση της μουσικής, την εξαγωγή και τη σχεδίαση σχεσιακών ηχητικών χαρακτηριστικών, ή την ανάπτυξη για μια καινοτόμα διεπαφή χρήστη. Η έρευνα στον τομέα MIR συγκεντρώνει ειδικούς με διαφορά ερευνητικά ενδιαφέροντα σε σχέση με την πληροφορική, την ανάκτηση πληροφορίας, τη μηχανική του ήχου, τη μουσικολογία, τη μουσική θεωρία καθώς και άλλους τομείς.

4.2 Εξαγωγή Χαρακτηριστικών

Για την δημιουργία συγκρίσιμων και προσβάσιμων μουσικών δεδομένων, το πρώτο βήμα σε όλες τις διεργασίες μουσικής επεξεργασίας είναι η εξαγωγή των κατάλληλων χαρακτηριστικών (feature extraction). Αυτά μπορούν να αποδώσουν σχετικά κλειδιά, καθώς απορρίπτουν μη σχετικές λεπτομέρειες. Εδώ η έννοια της ομοιότητας είναι κρίσιμης σημασίας για την σχεδίαση των ηχητικών χαρακτηριστικών. Για παράδειγμα για τη διάκριση τμημάτων μεταξύ ομιλίας και μουσικής, ένα υποψήφιο χαρακτηριστικό είναι η απόκλιση της ενέργειας του σήματος, καθώς αυτό το χαρακτηριστικό έχει μια φυσική ερμηνεία, η όποια έχει καλή εφαρμογή σε συγκεκριμένη διεργασία της κατηγοριοποίησης.

Σκοπός είναι η εξαγωγή ενός συνόλου χαρακτηριστικών από το δοθέν σύνολο δεδομένων. Τα χαρακτηριστικά αυτά πρέπει να περιέχουν επαρκή πληροφορία για τις επιθυμητές λεπτομέρειες της αρχικής προέλευσης των δεδομένων. Για την επίτευξη του στόχου πρέπει να γνωρίζουμε τον τομέα της εφαρμογής ώστε να επιλέξουμε τα κατάλληλα χαρακτηριστικά. Επομένως επιλογή των χαρακτηριστικών γίνεται βάσει του ζητούμενου αποτελέσματος. Για παράδειγμα αν έχουμε περιορισμένο χώρο αποθήκευσης ή επιθυμούμε ταχύτερη απόδοση του αλγορίθμου, τα χαρακτηριστικά επιλέγονται μέσω της μείωση των δεδομένων.

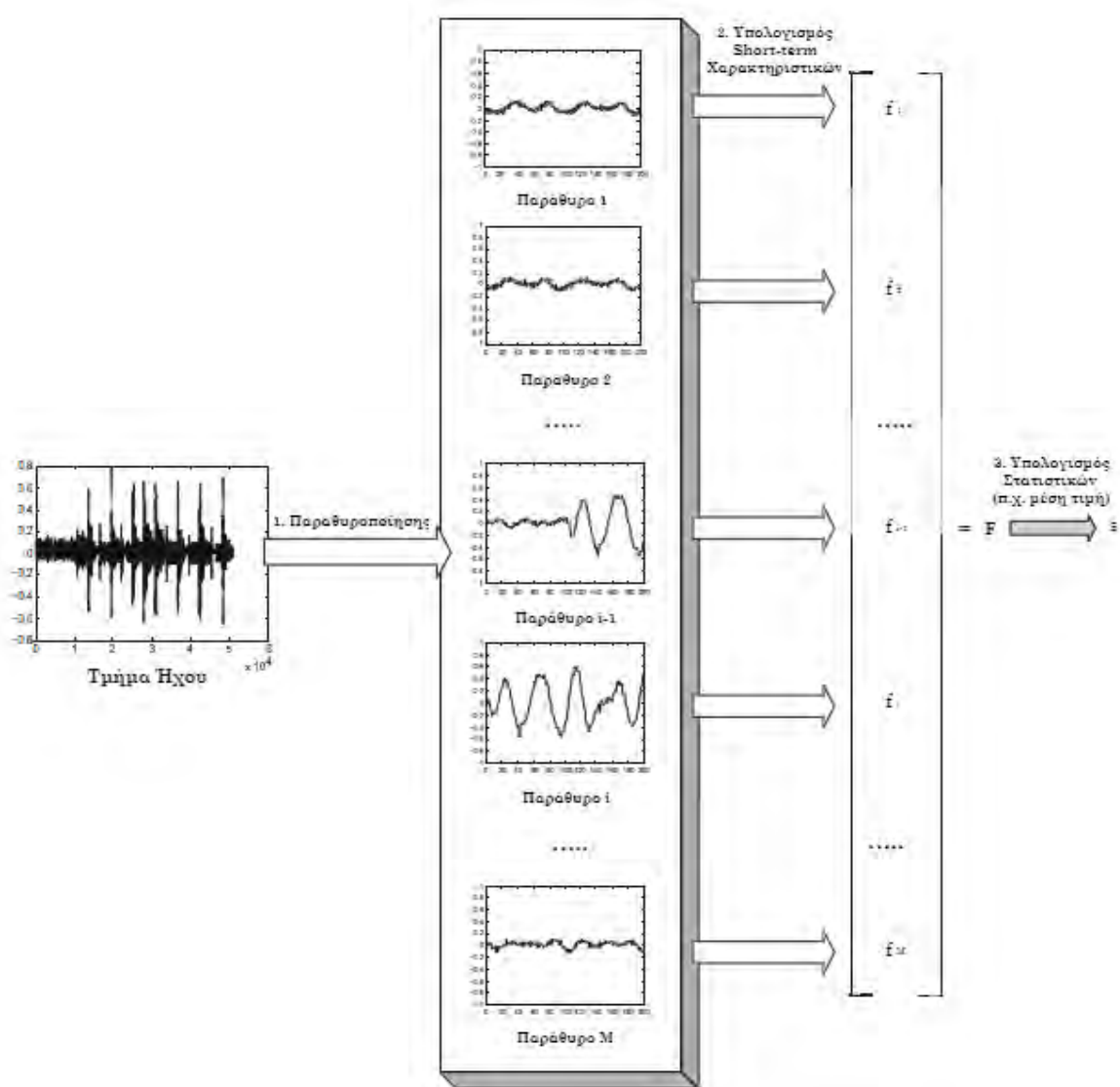
Στις παραγράφους που ακολουθούν περιγράφονται ορισμένες από τις μεθόδους εξαγωγής χαρακτηριστικών. Από αυτές υλοποιήθηκαν οι δύο για την παρόν εργασία, βραχυπρόθεσμη και ενδιάμεση εξαγωγή χαρακτηριστικών για τον ήχο και εξαγωγή χαρακτηριστικών κειμένου με “σώμα κειμένου” για τους στίχους. Επιπλέον αναφέρονται τρεις ακόμα μεθοδολογίες που είναι εξίσου διαδεδομένες στην εξαγωγή χαρακτηριστικών του ήχου, για τονικά, χρωματικά και CENS χαρακτηριστικά.

4.2.1 Εξαγωγή Short-Term Και Mid-Term Χαρακτηριστικών

Στις περισσότερες μεθόδους ανάλυσης και επεξεργασία ήχου, το σήμα διαιρείται σε βραχυπρόθεσμα παράθυρα (short-term frames). Το σήμα διασπάζεται σε επικαλυπτόμενα παράθυρα και μια ομάδα χαρακτηριστικών υπολογίζεται ανά παράθυρο. Αυτή του είδους η διεργασία αναπαράγει μια συχνότητα (F), από πίνακες χαρακτηριστικών για κάθε ηχητικό σήμα. Η διάσταση του πίνακα των χαρακτηριστικών εξαρτάται από την φύση των χαρακτηριστικών. Δεν είναι ασυνήθιστο να χρησιμοποιούμε μόνο μια διάσταση χαρακτηριστικών, όπως είναι ενέργεια του σήματος, ωστόσο στις πιο προηγμένες εφαρμογές ηχητικής ανάλυσης γίνεται εξαγωγή πολλαπλών χαρακτηριστικών τα οποία και συνδυάζονται σε πίνακες χαρακτηριστικών αυξημένης διάστασης. Οι εξαγόμενες συχνότητες των πινάκων των χαρακτηριστικών μπορούν να χρησιμοποιηθούν για επακόλουθες επεξεργασίες ή αναλύσεις των ηχητικών δεδομένων.

Εξίσου διαδεδομένη τεχνική είναι η επεξεργασία της ακολουθίας των χαρακτηριστικών με ενδιάμεσα παράθυρα (mid-term frames). Σύμφωνα με αυτόν τον τύπο επεξεργασίας το ηχητικό σήμα διαιρείται πρώτα σε ενδιάμεσα τμήματα. Έπειτα για κάθε τμήμα ακολουθεί βραχυπρόθεσμη επεξεργασία. Στο επόμενο βήμα η εξαγόμενη συχνότητα των χαρακτηριστικών (F) που δημιουργήθηκε από την ενδιάμεση διαίρεση, χρησιμοποιείται για τον υπολογισμό των στατιστικών των χαρακτηριστικών, όπως για

παράδειγμα η Μέση Τιμή Αλλαγής Πρόσημου (Average Zero-Crossing Rate). Στο τέλος το ενδιάμεσο τμήμα θα εκπροσωπείται από ένα σύνολο στατιστικών που αντιστοιχεί σε μια βραχυπρόθεσμη ακολουθία χαρακτηριστικών. Κατά την διάρκεια της ενδιάμεσης επεξεργασίας υποθέτουμε ότι τα τμήματα αυτά παρουσιάζουν ομογενής συμπεριφορές όσον αφορά τον τύπο του ηχητικού και συνεπώς εκφράζει λογική ώστε να προχωρήσει στην εξαγωγή των στατιστικών βάση του τμήματος. Στην πράξη η διάρκεια των ενδιάμεσων παραθύρων τυπικά κυμαίνεται σε απόσταση ενός με δέκα δευτερόλεπτων, το οποίο εξαρτάται από το πεδίο εφαρμογής.



Σχήμα 4. Διαδικασία εξαγωγής Mid-term χαρακτηριστικών.

4.2.2 Εξαγωγή Τονικών Χαρακτηριστικών

Μια άλλη μέθοδος εξαγωγής είναι μέσω τονικών χαρακτηριστικών (pitch features). Κάθε μουσική νότα αντιστοιχίζεται με ένα MIDI τόνο (p). Επιπροσθέτως η σχετική συχνότητα (associated frequency) μιας νότας αναφέρεται και ως κεντρική της συχνότητα. Συνοψίζοντας θεωρούμε ότι αναφερόμαστε σε νότες εμβέλειας A0 ή Λα 0ης οκτάβας ($p = 21$) έως C8 ή Ντο 8ης οκτάβας ($p = 108$) με τις αντίστοιχες τονικότητες, οι οποίες αντιστοιχούν σε ένα μέσω πιάνο. Για κάθε τόνο, σχεδιάζουμε ένα ζωνοπερατό φίλτρο (bandpass filter), το οποίο προσπελαύνει όλες τις συχνότητες γύρω από την κεντρική συχνότητα καθώς απορρίπτει όλες τις άλλες συχνότητες.

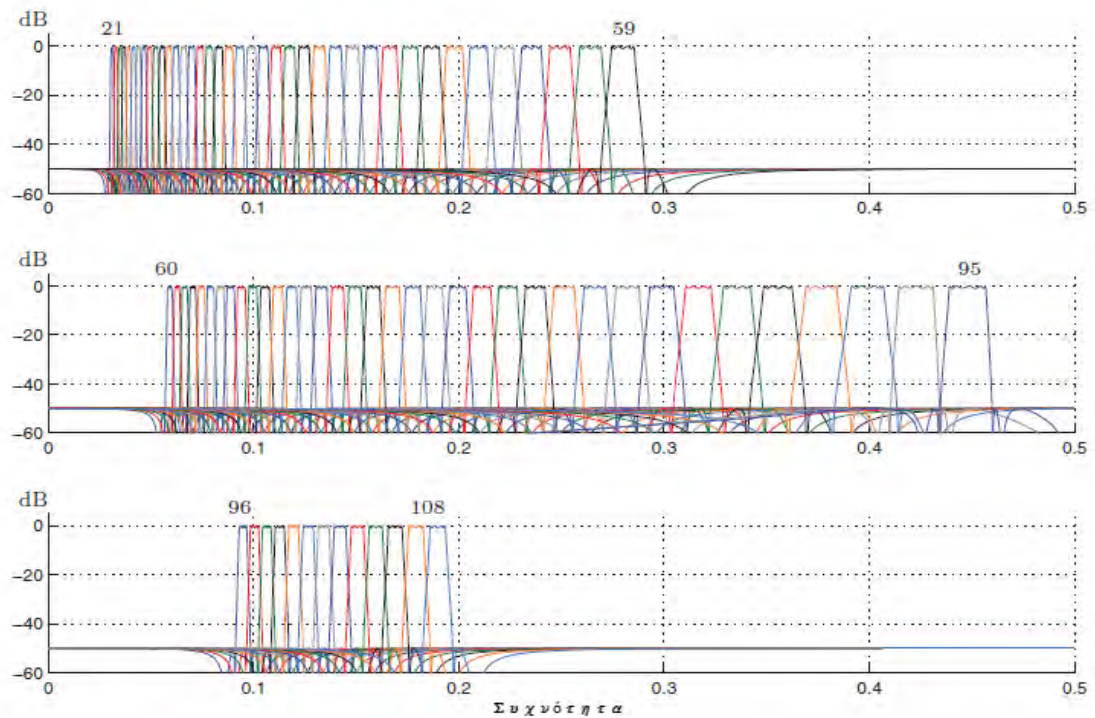
Οι απαιτήσεις του φίλτρου που αναφέραμε παραπάνω είναι αυστηρές ως προς τον διαχωρισμό των γειτονικών νοτών, η ζωνοπερατότητα των φίλτρων πρέπει να είναι περιορισμένη, το κατώφλι αποκοπής πρέπει να είναι ευκρινές, η απόρριψη πρέπει να είναι υψηλή. Επιπλέον η διάταξη του φίλτρου θα πρέπει να είναι συμπαγής για αποτελεσματικό υπολογισμό. Ο αριθμός της MIDI νότας, οι νότες ισοσκελής κλίμακας (equal-tempered scale), εξαρτάται από τις κεντρικές συχνότητες και συνδέεται με λογαριθμικό τύπο. Έπεται ότι το εύρος ζώνης ενός φίλτρου που συσχετίζεται με χαμηλό τόνο θα πρέπει να είναι μικρότερο από αυτό ενός υψηλού τόνου. Ωστόσο, το πολύ μικρό εύρος ζώνης και η σύντομη μετάβαση ίσως οδηγήσει σε αριθμητικά προβλήματα στην σχεδίαση του φίλτρου καθώς και στην επεξεργασία του φιλτραρίσματος. Για την ελάφρυνση των απαιτήσεων του φίλτρου, ένας τρόπος είναι η χρήση διαφορετικών τιμών δειγμάτων. Προσαρμόζοντας τις τιμές των δειγμάτων εκτιμάται ότι ο χρόνος ανάλυσης μειώνεται με φυσικό τρόπο για χαμηλότερες συχνότητες. Αντίστοιχα, η μείωση των τιμών των δειγμάτων προκαλεί επιτάχυνση των υπολογισμών. Η διχοτόμηση των τόνων και η κατανομή των τιμών δειγμάτων αποτελεί ένα συμβιβασμό μεταξύ ανάλυσης, αριθμητικής ευστάθειας και υπολογιστικής πολυπλοκότητας.

Λόγω των ακριβή λεπτομερών του ορίου αποκοπής, χρησιμοποιούνται ελλειψοειδή φίλτρα. Για το συμβιβασμό της αριθμητικής σταθερότητας στη σχεδίαση, στην διάταξη και στα χαρακτηριστικά του φίλτρου, κάθε φίλτρο είναι υλοποιημένο με ελλειψοειδές φίλτρο 8^{ου} βαθμού, με 1dB ζωνοπερατότητα και 50 dB κατώφλι για την απόρριψη συχνοτήτων στην ζώνη διακοπής. Το εύρος ζώνης του φίλτρου συνήθως καθορίζεται μέσα από αυτό που αποκαλείτε “παράγοντα ποιότητας” (quality factor ή Q factor), ο οποίος περιγράφει την τιμή της κεντρικής συχνότητας του εύρους ζώνης. Για τον διαχωρισμό των νοτών, χρησιμοποιεί το Q factor και μια ζώνη μετάδοσης η όποια έχει το μισό πλάτος από αυτή του ζωνοπερατού. Έως τώρα οι προδιαγραφές του φίλτρου

δίνονται σε απόλυτες τιμές εφόσον αναφερόμαστε σε σήματα συνεχούς χρόνου (CT). Για την απόκτηση των προδιαγραφών του φίλτρου όσον αφορά ένα συγκεκριμένο δείγμα τιμών, πρέπει να διαχωρίσουμε τις οντότητες με βάση αυτό.

Γενικά, ένας πίνακας ζωνοπερατών φίλτρων που διαχωρίζει το σήμα εισόδου σε πολλαπλά τμήματα αναφέρεται ως “τράπεζα φίλτρων” (filter bank). Για την λήψη της εξόδου εφαρμόζεται ένα φίλτρο στο σήμα εισόδου, το οποίο ονομάζεται υποζώνη (subband) του σήματος εισόδου. Εξ ορισμού, μια τράπεζα φίλτρων πολλαπλών τιμών (multirate) περιλαμβάνει πολλαπλές τιμές δειγμάτων. Τα φίλτρα που προαναφέραμε από την τράπεζα φίλτρων έχουν το ίδιο Q factor.

Προτού εφαρμόσουμε την τράπεζα φίλτρων, πρώτα πρέπει να παρέχουμε στο σήμα εισόδου τις απαραίτητες τιμές δειγμάτων. Αφού εφαρμόσουμε ένα χαμηλοπερατό φίλτρο εξομάλυνσης με κάποια όριο αποκοπής, το σήμα μειώνεται ως προς το δείγμα (downsampling) με βάση ενός συντελεστή ώστε να προκύψει ένα διακριτό σήμα (DT) έστω x_1 . Ακολουθώντας την ίδια διαδικασία το οδηγεί σε ένα διακριτό σήμα x_2 . Για να αποκτήσουμε μια τονική αναπαράσταση του αρχικού σήματος, κάθε φίλτρο στην τράπεζα φίλτρων εφαρμόζεται σε ένα από τα διακριτά σήματα x_0 (σήμα εισόδου), x_1 και x_2 , σύμφωνα με την απαιτούμενη τιμή δείγματος. Η ομαδική καθυστέρηση διορθώνεται χρησιμοποιώντας προς τα εμπρός και προς τα πίσω φιλτράρισμα.



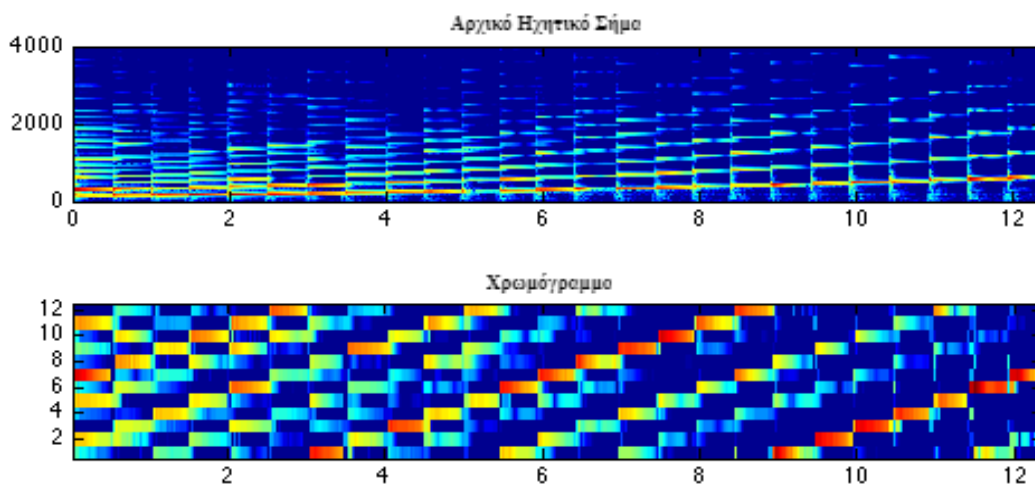
Σχήμα 5. Τα τρία διαγράμματα περιγράφουν τις φάσεις μείωσης των δειγμάτων των σημάτων x_0 , x_1 και x_2 .

4.2.3 Εξαγωγή Χρωματικών και CENS Χαρακτηριστικών

Η ανθρώπινη αντίληψη του τόνου είναι περιοδική ως προς την αίσθηση ότι δυο τόνοι μπορούν να θεωρηθούν ισοδύναμοι με κάποια χρωματική αναπαράσταση εάν διαφέρουν κατά μια οκτάβα. Βάση αυτού του σημείου παρατήρησης, ένας τόνος μπορεί να διαχωριστεί σε δυο τμήματα, το τονικό ύψος (tone height) και το χρώμα (chroma). Στα ακόλουθα θεωρούμε τις μουσικές νότες ισοδύναμης κλίμακας. Έχοντας ως συμβολισμό το MIDI πρότυπο, το τονικό ύψος αναφέρεται στον αριθμό οκτάβας και το χρώμα στο αντίστοιχο τονικό χαρακτηριστικό γραφής το οποίο περιέχεται στη λίστα από νότες ενός απλού πιάνου {C,C#,D, . . . ,B}. Σε αυτό θα πρέπει να επισημάνουμε ότι στις νότες ισοδύναμης κλίμακας, διαφορετικοί τόνοι όπως το C# ή ντο δίεση και το Db ή ρε με ύφεση b αναφέρονται στο ίδιο χρώμα. Η τονική κλάση (pitch class) ορίζεται ως μια σειρά από τόνους που έχουν ως χαρακτηριστικό τους το ίδιο χρώμα.

Η χρωματική αναπαράσταση μπορεί να επιτευχθεί από μια τονική αναπαράσταση προσθέτοντας υποζώνες που αντιστοιχούν στην ίδια κλάση. Πιο συγκριμένα, αρχικά πρέπει να διαιρέσουμε το αρχικό σήμα σε υποζώνες και να υπολογιστεί για κάθε υποζώνη το STMSP (Short-Time Mean-Square Power) και αμέσως μετά προσθέτουμε όλα τα STMSPs που ανήκουν στην ίδια τονική κλάση. Το αποτέλεσμα από ένα STMSP

αντιστοιχεί σε ένα χρώμα. Για παράδειγμα το χρώμα C, προέρχεται από το άθροισμα των τόνων C1, C2, ... C8 που αντιστοιχούν στην ίδια κλάση. Αυτό παράγει ένα πολυδιάστατο πίνακα όπου κάθε στοιχείο αντιστοιχεί σε ένα χρώμα. Η αναπαράσταση του αποτελέσματος αποκαλείται ως “χρωματική αναπαράσταση” του ηχητικού σήματος. Το ηχόχρωμα ενός ήχου σχετίζεται με την ενεργειακή κατανομή των αρμονικών. Επομένως κατά την ισοδυναμία της οκτάβας, τα χρωματικά χαρακτηριστικά αποδίδουν σε ένα μεγάλο βαθμό τις παραλλαγές του ηχοχρώματος. Επιπρόσθετα τα χρωματικά χαρακτηριστικά είναι υπεύθυνα για την σχέση οκτάβας σε μελωδία και σε αρμόνια σύμφωνα με τη διάκριση στη Δυτική μουσική. Αυτό είναι ιδανικό για τη μουσική ανάλυση όπου χαρακτηρίζεται από αρμονική μετάβαση.



Σχήμα 6. Το επάνω γράφημα είναι αναπαράσταση ενός ηχητικού σήματος όπου ο άξονας των x αντιπροσωπεύει τον χρόνο ενώ ο άξονας των y τη συχνότητα. Το κάτω γράφημα το οποίο ονομάζεται “χρωμόγραμμα” (chromagram) και είναι η χρωματική αναπαράσταση του αρχικού σήματος όπου επίσης ο άξονας των x αντιπροσωπεύει τον χρόνο ενώ ο άξονας των y αντιπροσωπεύει το χρώμα (chroma).

4.2.4 Εξαγωγή Χαρακτηριστικών από Κείμενο

Μια ιδιαίτερη και σημαντική μέθοδος στην εξαγωγή χαρακτηριστικών και ευρύτερα στην Μηχανική Μάθηση είναι η κατηγοριοποίηση κειμένου. Η κατηγοριοποίηση κειμένου στοχεύει στην εύρεση κοινών θεματικών στοιχείων, ώστε να κατατάξει την οποιαδήποτε λέξη που υπάρχει στο κείμενο σε μια ομάδα. Για να γίνει πιο αντιληπτή αυτή η ομαδοποίηση των λέξεων θα αναφερθούμε σε ένα απλό παράδειγμα. Έστω έχουμε τις λέξεις “ουτοπία”, “τοποθεσία” και “τοποθετώ”, οι οποίες έχουν ένα βασικό κοινό

χαρακτηριστικό γνώρισμα: είναι ομόριζες και προέχονται από την λέξη “τόπος”. Αυτός είναι ένας τρόπος να εντάξουμε στην ίδια ομάδα ένα σύνολο λέξεων, το οποίο μας βοηθάει στο διαχωρισμό και στην επεξεργασία της πληροφορίας μεγάλου εύρους κειμένων.

Ένα ισχυρό δομοστοιχείο στον τομέα της διαχείρισης μεγάλου όγκου κειμένων είναι το “σώμα κειμένου”(corpus). Το σώμα κειμένου είναι μια συλλογή από κείμενα, τα οποία έχουν γραφτεί ή έχουν διατυπωθεί προφορικά, αποτυπωμένα σε ηλεκτρονική μορφή. Μέσα από το σώμα κειμένου μπορούμε να λάβουμε πληροφορίες γλωσσολογικού περιεχομένου για το αρχικό κείμενο. Αυτή την πληροφορία εκμεταλλεύονται οι αλγόριθμοι Μηχανικής Μάθησης για τη δημιουργία στατιστικών μοντέλων. Για να εξάγουμε την χρήσιμη πληροφορία από το κείμενο απαραίτητη είναι η διαίρεση του σε υπομονάδες. Για παράδειγμα τέτοιου είδους υποδιαιρέσεις θα ήταν λέξεις και σημεία στίξης. Αυτή η διαδικασία ονομάζεται Διαμέριση (Segmentation) και Ανάθεση Λεκτικής Μονάδας (Tokenization).

5. Παλινδρόμηση

5.1 Το Στατιστικό Μοντέλο της Παλινδρόμησης

Στην στατιστική η παλινδρόμηση είναι ένα μοντέλο που ερευνά τη συσχέτιση μεταξύ μιας εξαρτώμενης μεταβλητής και μιας ή περισσότερων ανεξαρτήτων μεταβλητών με σκοπό την πρόβλεψη των τιμών της μίας, μέσω των τιμών της άλλης. Συνεπώς από την μια συναντάμε τις ανεξάρτητες ή ελεγχόμενες ή επεξηγηματικές μεταβλητές (independent, predictor, explanatory variables) που συνηθίζεται να απεικονίζονται στον άξονα των x των Καρτεσιανών συντεταγμένων. Από την άλλη έχουμε τις εξαρτημένες ή μεταβλητές απόκρισης (dependent, response variables) όπου αντίστοιχα συνηθίζουν να απεικονίζονται από τον άξονα y των Καρτεσιανών συντεταγμένων.

5.2 Ορισμοί του Μοντέλου Παλινδρόμησης

Ας ερμηνεύσουμε το μοντέλο σύμφωνα με ένα παράδειγμα των Schroeder, Sjoquist και Stephan (1986). Έστω δύο καταστάσεις “Όσο μεγαλύτερη είναι η δαπάνη ενός πολιτικού, στην προεκλογική καμπάνια, τόσο μεγαλύτερος θα είναι και ο αριθμός των ψήφων που θα λάβει.” και “Η Μαρία είναι ψηλότερη από την Ιωάννα.” όπου εκφράζουν διαφορετικού τύπους σχέσεων. Στην πρώτη εκφράζεται η σχέση εξάρτησης του ποσοστού των ψήφων που θα λάβει ο υποψήφιος σε συνάρτηση με το δαπανηθέν προεκλογικό ποσό ενώ στη δεύτερη κανένα αίτιο δεν υποδεικνύεται. Μια από τις δραστηριότητες των ερευνητών, αυτού του τομέα, είναι να εξετάζουν την ορθότητα ή την ανακρίβεια των υποθετικών συναρτησιακών σχέσεων που αποκαλούνται υποθέσεις (hypothesis) ή θεωρίες (theories). Ένα από τα εργαλεία του ελέγχου της υπόθεσης είναι η Παλινδρόμηση.

Ο Γιαννακόπουλος, Θ. (2015) δίνει ορισμό για την παλινδρόμηση ως την διεργασία που εκτιμά την τιμή μιας άγνωστης μεταβλητής, δεδομένου του αντιστοίχου πίνακα χαρακτηριστικών. Δίνοντας με αυτόν τον τρόπο έναν πιο εξειδικευμένο ορισμό τοποθετώντας την εξαγωγή χαρακτηριστικών και συναισθημάτων, μέσα από αναπαράσταση σε διάσταση, ενός συνόλου δεδομένων στο ρόλο της εξαρτώμενης και ανεξάρτητης μεταβλητής. Αναφέρει επίσης ότι η παλινδρόμηση χρησιμοποιείται για τη χαρτογράφηση των χαρακτηριστικών (ηχητικών, κειμένου κλπ.) σε πραγματικών τιμών μεταβλητές.

5.3 Παλινδρόμηση με Μηχανές Διανυσμάτων Υποστήριξης

Οι Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machines/SVM) ανήκουν στην περιοχή της Επιβλεπόμενης Μάθησης (Supervised Learning) και έχουν ως στόχο τη δημιουργία συστημάτων λήψεων αποφάσεων τα οποία προσπαθούν να προβλέψουν νέες τιμές. Σύμφωνα με τον Francesco Parrella (2007) η τεχνική αυτή υποδιαιρείται σε δυο στάδια.

Στάδια 1. Μάθηση: Που αποτελείται από ένα εκπαιδευόμενο SVM.

Στάδια 2. Πρόβλεψη: Που νέα δείγματα εισάγονται, για τα οποία δεν γνωρίζουμε τα αποτελέσματά τους. Τα αποτελέσματα θα εκτιμηθούν σύμφωνα με την εκπαίδευση που έγινε στο προηγούμενο στάδιο.

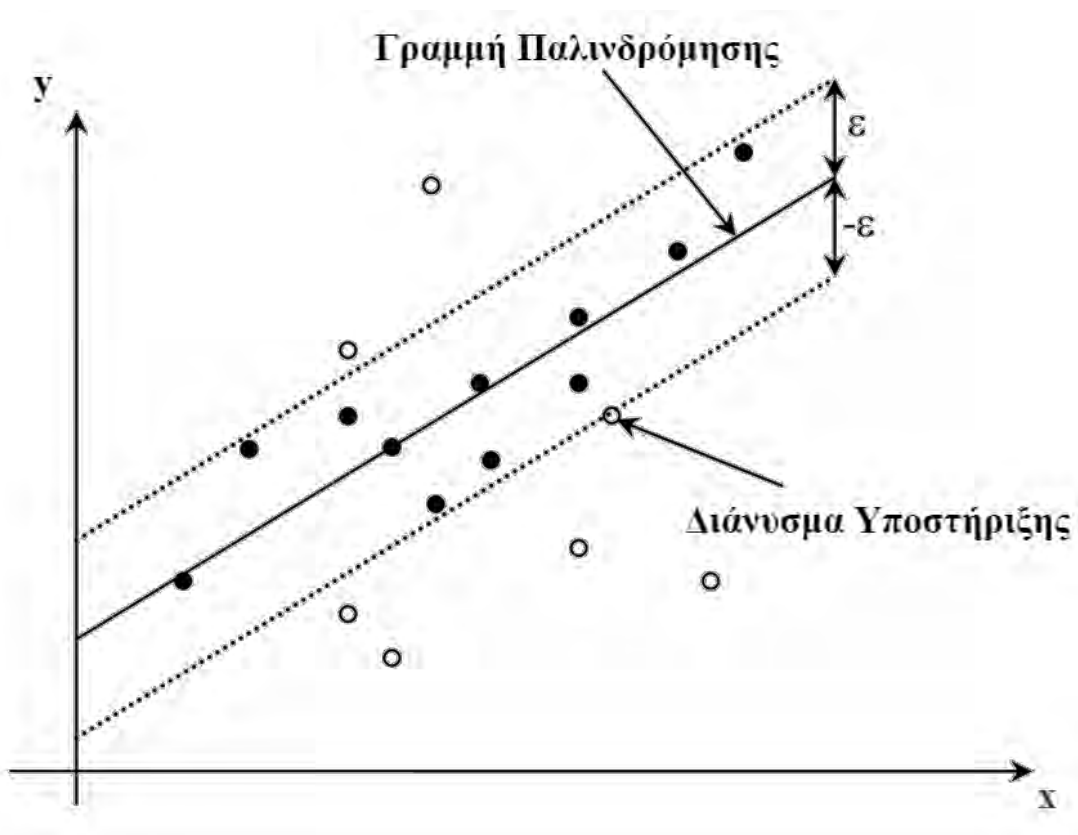


Σχήμα 7. Αναπαράσταση του SVM μοντέλου σε υψηλό επίπεδο.

Για να γίνουμε πιο σαφείς στο θέμα της ορολογίας το κομμάτι των Μηχανών Διανυσμάτων Υποστήριξης που ασχολείται με την παλινδρόμηση αποκαλείται Παλινδρόμησης Διανυσμάτων Υποστήριξης (Support Vector Regression/SVR). Διαχωρίζοντας με αυτόν τον τρόπο ότι με τον αρχικό όρο SVM αναφερόμαστε στην Ομαδοποίηση.

5.3.1 Μαθηματική Προσέγγιση

Όπως αναφέρθηκε στην αρχή του κεφαλαίου η αναπαράσταση των τιμών αποτυπώνεται στο Καρτεσιανό σύστημα συντεταγμένων. Συνεπώς αναφερόμαστε σε ζεύγη τιμών (x, y) όπου το x είναι ένας πίνακας που αποτελείται από πραγματικούς αριθμούς ενώ η έξοδος μπορεί είναι και αυτή πραγματικό αριθμός ή λογική τιμή Μπούλ μορφότυπου (Boolean) όπως για παράδειγμα σωστό/λάθος ή 1/-1. Η πρώτη περίπτωση είναι που αφορά την Παλινδρόμηση ενώ η δεύτερη την Ομαδοποίηση. Ουσιαστικός στόχος των διανυσμάτων είναι να εντοπίσουν το ευθύγραμμο τμήμα που διαχωρίζει το δείγμα με το καλύτερο δυνατό τρόπο, με σκοπό να είναι διακριτή η διαφορετικότητα τους (βάση κάποιου χαρακτηριστικού γνωρίσματος). Το διαχωριστικό αυτό ευθύγραμμο τμήμα ονομάζεται υπερεπίπεδο (hyperplane). Όσον αφορά την αναπαράσταση των τιμών στο χώρο, τα αποτελέσματα από την μάθηση μπορούν να συσχετισθούν με μια πραγματική συνάρτηση και το Διάνυσμα Υποστήριξης αναπαρίσταται από την συνάρτηση όπου παρεμβάλει, με την καλύτερη δυνατότητα, την σειρά των σημείων αυτών.



Σχήμα 8.Γραφική αναπαράσταση SVR.

Ένα ποσοστό από το δείγμα χρησιμοποιείται για τη σύνθεση των Διανυσμάτων Υποστήριξης και καλείται Σύνολο Εκπαίδευσης (Training Set) ενώ το υπόλοιπο δείγμα χρησιμοποιείται για την πρόβλεψη και καλείται Σύνολο Ελέγχου (Test Set).

Σε αυτό το σημείο είναι σημαντικό να επισημάνουμε την διαφορά των δύο μοντέλων SVM και SVR. Όπως προαναφέρθηκε η διαφορά είναι στην τιμή εξόδου το SVR επιλύει το πρόβλημα της πρόβλεψης με πραγματικό αριθμό. Αυτό επιτυγχάνεται με θέσπιση ορίων ϵ , για την προσέγγιση του SVM, όπως φαίνεται και στο σχήμα 10. Παρόλα αυτά ο βασικός στόχος παραμένει ίδιος: μείωση του σφάλματος και εξατομίκευση του υπερεπίπεδου (hyperplane) με την όποια μεγιστοποιούνται τα όρια, κάτι το οποίο είναι αποδεκτό.

6. Μεθοδολογία και Εργαλεία

6.1 Εργαλεία που Χρησιμοποιήθηκαν για την Υλοποίηση της Εργασίας

Η εργασία υλοποιήθηκε με τη γλώσσα python, πιο συγκεκριμένα την έκδοση 2.7 καθώς και με την αξιοποίηση κάποιων δομοστοιχείων/βιβλιοθηκών της. Οι κυρίες βιβλιοθήκες που χρησιμοποιήθηκαν είναι το youtube_dl για την συλλογή των δεδομένων στον κομμάτι που αφορά τα ηχητικά δεδομένα, για τη συλλογή των στίχων η PyLyrics και τέλος για την καταγραφή των μεταβλητών μέτρησης των συναισθημάτων (Valance, Danceability και Energy) το spotipy. Επιπλέον για την επεξεργασία και αποθήκευση (σε αρχεία μορφής .tsv ή .csv) των δεδομένων που συλλέχτηκαν χρησιμοποιήθηκαν τα Pandas. Για την εξαγωγή των χαρακτηριστικών (Feature Extraction), την παλινδρόμηση (Regression) και την εκπαίδευση του δικτύου και ευρύτερα την επεξεργασία των ηχητικών δεδομένων χρησιμοποιήθηκε η pyAudioAnalysis (η έκδοση από το github). Αντίστοιχα για την εξαγωγή χαρακτηριστικών στο τμήμα που αφορά τους στίχους χρησιμοποιήθηκε το genism ενώ η εκπαίδευση του συστήματος έγινε με κώδικα της pyAudioAnalysis.

6.1.1 youtube-dl

Το youtube_dl παρέχει την δυνατότητα λήψης βίντεο από ιστοσελίδες όπως το YouTube.com με προσαρμογή της κωδικοποίησης. Για την παρόν εργασία χρησιμοποιήθηκε το mp3 μορφότυπο. Αρχικά συγκρίνει τη δοθείσα ηλεκτρονική διεύθυνση με τα υποπρόγραμμα του πακέτου εξαγωγής στο φάκελο extractor, όπου το καθένα έχει συγκριμένους κανόνες που αναλογούν στις αντίστοιχες ιστοσελίδες ώστε να μπορεί να εξάγει το περιεχόμενο των ηλεκτρονικών διευθύνσεων μέσα από αυτές. Όταν ο εξαγωγέας επιστρέψει την ηλεκτρονική διεύθυνση μαζί με κάποια άλλα μεταδιδόμενα, καλείται η συνάρτηση get_suitable_downloader από το αρχείο downloader έχοντας ως παραμέτρους το περιεχόμενο των ηλεκτρονικών διευθύνσεων που επέστρεψε ο εξαγωγέας για να καθοριστεί το πρωτόκολλο που θα χρησιμοποιηθεί κατά την διαδικασία λήψης. Η εφαρμογή του κάθε πρωτοκόλλου προεκτείνεται μέσω της κλάσης FileDownloader, η οποία παρέχει την διεπαφή για την λήψη του αρχείου. Ανάλογα με τις επιλογές το ληφθέν αρχείο θα χρειαστεί και τις ανάλογες εκ των υστέρων διεργασίες, για παράδειγμα η αποσυμπίεση του ήχου από το βίντεο για να παραχθεί ξεχωριστό αρχείο ήχου. Τα υποπρογράμματα τα οποία είναι υπεύθυνα τέτοιου είδους διεργασίες βρίσκονται στο φάκελο postprocessor.

6.1.2 PyLyrics

Η PyLyrics με την χρήση πολλών αφηρημένων επιπέδων παρέχει μουσικά δεδομένα, όπως στίχους, μέσα από την ιστοσελίδα lyrics.wikia.com. Τρεις βασικές κλάσεις ερμηνεύουν τις περισσότερες από τις τιμές που επιστρέφει η PyLyrics ἡ κλάση Track που επιστρέφει τιμές σε αναζητήσεις μουσικών κομματιών κάποιου δίσκου και περιέχει μεθόδους για την εξαγωγή των στίχων, η κλάση Album που επιστρέφει τιμές ύστερα από αναζητήσεις με περιεχόμενο τον δίσκο ενός καλλιτέχνη, η οποία περιέχει μεθόδους για την εξαγωγή μουσικών κομματιών και τέλος η κλάση Artist που επιστρέφει τιμές σε αναζητήσεις για τον καλλιτέχνη, ενώ περιέχει μεθόδους για την εξαγωγή μουσικών δίσκων.

6.1.3 Spotipy

Το spotipy είναι μια βιβλιοθήκη της python που παρέχει διασύνδεση με το διαδικτυακό API της εφαρμογή Spotify. Παρέχει πλήρη πρόσβαση στα μουσικά δεδομένα που παρέχει η πλατφόρμα του Spotify καθώς και επιπλέον όγκο πληροφορίας. Η πρόσβαση μπορεί να γίνει με ή χωρίς εξουσιοδότηση. Οι κλάσεις και οι συναρτήσεις για την εξουσιοδοτημένη πρόσβαση βρίσκονται στο αρχείο oauth2. Για οποιαδήποτε αίτημα αναζήτησης καλλιτέχνη, δίσκων και οτιδήποτε παρεμφερές , το αρχείο client παρέχει συναρτήσεις όπως η search, album και artist με τα ανάλογα ορίσματα. Για την επίτευξη των στόχων της εφαρμογής υλοποιήθηκε η συνάρτηση search και γίνεται αναζήτηση με βάση τον τίτλο του κομματιού (ως τίτλο αναφερόμαστε στον τίτλο που παρέχεται στο βίντεο από τον ιστότοπο του Youtube αφαιρώντας πληροφορίες όπως καλλιτέχνες που συμμετέχουν στον κομμάτι και διατηρώντας μόνο τον κύριο ερμηνευτή και την ονομασία του κομματιού που ερμηνεύει.

6.1.4 Pandas

Τα Pandas είναι μια βιβλιοθήκη που προσφέρει υψηλής απόδοσης και ευκολία στην χρήση εργαλείων δόμησης και ανάλυσης δεδομένων. Τα δυο βασικά δομικά στοιχεία της βιβλιοθήκης είναι οι κλάσεις των ακολουθιών (Series\Μονοδιάστατο περιβάλλον) και τα πλαίσια δεδομένων (DataFrame\Διδιάστατο περιβάλλον), τα οποία διαχωρίζουν την πλειοψηφία των τυπικών περιπτώσεων στα οικονομικά, στη στατιστική, στις κοινωνικές

επιστήμες και διάφορα τμήματα της μηχανολογίας. Μεταποίηση Μεγέθους (Προσθήκη \ Αφαίρεση στύλων ή γραμμών), Στοίχιση Δεδομένων (Τα δεδομένα ενός πλαισίου μπορούν να διαταχθούν βάση των ετικετών), Εύκολης Μετατροπής (Γρήγορη μετατροπή μπορεί να προκύψει ακόμα και ανά στήλη) είναι ορισμένα από τα εργαλεία που παρέχει η βιβλιοθήκη των Pandas. Η ευελιξία των Pandas στο να διαχωρίζονται μεγάλου όγκου πληροφορίες οι οποίες και διατηρούνται σε πολυδιάστατους πίνακες που επιτρέπουν την ανάγνωση και την αποθήκευση σε αρχεία μορφότυπου .tsv (εικόνα 4).



Εικόνα 4. Αρχεία με το μορφότυπο tsv όπου βρίσκονται αποθηκευμένες οι πληροφορίες αρχειοθετημένες σε πίνακες μονοδιάστατους και διδιάστατους αντίστοιχα.

6.1.5 pyAudioAnalysis

Η pyAudioAnalysis είναι μια βιβλιοθήκη που παρέχει ένα ευρύ φάσμα λειτουργικοτήτων σχετικές με τον ήχο, εστιάζοντας κατά κύριο λόγο στην Εξαγωγή των Χαρακτηριστικών (Feature Extraction), στην Κατηγοριοποίηση (Classification), την Παλινδρόμηση (Regression), τη Διαμέριση (Segmentation) και σε θέματα Απεικόνισης (Visualization). Υπάρχουν δυο στάδια εξαγωγής χαρακτηριστικών στην μεθοδολογία του ήχου. Το πρώτο στάδιο ονομάζεται Βραχυπρόθεσμη (Short-term) Εξαγωγή Χαρακτηριστικών. Εκτελείται μέσω της συνάρτησης `stFeatureExtraction` στο φάκελο `audioFeatureExtraction` όπου διαχωρίζει το εισαγόμενο σήμα σε βραχυπρόθεσμα πλαίσια και υπολογίζει των αριθμό των χαρακτηριστικών για κάθε πλαίσιο. Κάθε βραχυπρόθεσμο πλαίσιο έχει μέγεθος 20 με 100 ms. Από αυτή την διαδικασία παράγεται μια ακολουθία βραχυπρόθεσμων πινάκων για ολόκληρο το σήμα. Το δεύτερο στάδιο ονομάζεται Ενδιάμεση (Mid-term) Εξαγωγή Χαρακτηριστικών. Μέσω της συνάρτησης `mtFeatureExtraction` που όμοια βρίσκεται στο φάκελο `audioFeatureExtraction`, το σήμα χωρίζεται σε ενδιάμεσα πλαίσια. Τόσο η Βραχυπρόθεσμη όσο και η Ενδιάμεση Εξαγωγή μπορεί να υλοποιηθεί επικαλυπτόμενα ή μη επικαλυπτόμενα (overlapping ή non-overlapping αντίστοιχα), δηλαδή το βήμα του

παράθυρου είναι μικρότερο ή ίσο (αντίστοιχα) σε σχέση με το μέγεθος του. Για κάθε τμήμα που δημιουργήθηκε από το πρώτο στάδιο της βραχυπρόθεσμης εξαγωγής και αντίστοιχα για τα χαρακτηριστικά του τμήματος της ενδιάμεσης εξαγωγής, χρησιμοποιούνται για τον υπολογισμό στατιστικών για χαρακτηριστικά. Ο παρακάτω πίνακας ερμηνεύει ένα ένα τα χαρακτηριστικά.

Πίνακας 1.Κάθε Βραχυπρόθεσμο πλαίσιο αντιπροσωπεύετε από ένα διάνυσμα 34 χαρακτηριστικών όπως παρουσιάζονται στον παρακάτω πίνακα.

Αριθμός Χαρακτηριστικού	Όνομα Χαρακτηριστικού	Περιγραφή
1	Ρυθμός Αλλαγής Προσήμου (Zero Crossing Rate)	Ο ρυθμό μεταβολή πρόσημου του σήματος την χρονική περίοδο της διάρκειας ενός συγκεκριμένου πλαισίου.
2	Ενέργεια (Energy)	Το άθροισμα των τετραγώνων των τιμών του σήματος, κανονικοποιημένο από το αντίστοιχο μήκος πλαισίου.
3	Εντροπία της Ενέργειας (Entropy of Energy)	Η εντροπία των υπό-πλαισίων την κανονικοποιημένων ενεργειών. Μπορεί να ερμηνευτεί ως μέτρηση των αποτόμων αλλαγών.
4	Κεντρικό Φάσμα (Spectral Centroid)	Το κέντρο της βαρύτητας του φάσματος.
5	Διευρυμένο Φάσμα (Spectral Spread)	Η δεύτερη κεντρική ροπή του φάσματος.

Αριθμός Χαρακτηριστικού	Όνομα Χαρακτηριστικού	Περιγραφή
6	Εντροπία Φάσματος (Spectral Entropy)	Η εντροπία των κανονικοποιημένων ενεργειών για μια ομάδα υπό-πλαισίων.
7	Φασματική Διακύμανση (Spectral Flux)	Η τετραγωνική διαφορά των κανονικοποιημένων μεγεθών του φάσματος των δυο διαδοχικών πλαισίων.
8	Φασματική Αναίρεση (Spectral Rolloff)	Η συχνότητα κάτω από την οποία το 90% του μεγέθους κατανομής του φάσματος έχει συσσωρευθεί.
9-21	Συντελεστές Συχνότητας με Σάφματα Mel (Mel Frequency Cepstral Coefficients/MFCCs)	Προκύπτει από αναπαράσταση σαφμάτων όπου οι ζώνες συχνοτήτων δεν είναι γραμμικές αλλά διαμοιρασμένες σύμφωνα την κλίμακα του Mel
22-33	Χρωματικός Πινάκας (Chroma Vector)	Αναπαράσταση δώδεκα στοιχείων της φασματικής ενέργειας, Αναπαριστούν τις δώδεκα ισοσυγκερασμένες κλάσεις τόνων δυτικού τύπου μουσικής (ημιτονική διάταξη).
34	Χρωματική Απόκλιση (Chroma Deviation)	Η σταθερή απόκλιση των δώδεκα χρωματικών συντελεστών.

Το τμήμα που είναι υπεύθυνο για την λειτουργικότητα της κατηγοριοποίησης δημιουργήθηκε με σκοπό να εκπαιδεύσει και με την χρήση μηχανικών μοντέλων να κατηγοριοποιήσει ένα άγνωστο ηχητικό κομμάτι , σε μια ομάδα προκαθορισμένων

Γ. Τσαμαδιάς

κλάσεων. Η βιβλιοθήκη υλοποιεί τις ακόλουθες μορφές κατηγοριοποίησης: κ Πλησιέστερος Γείτονας (k Nearest Neighbor/kNN), Μηχανής των Διανυσμάτων Υποστήριξης (Support Vector Machines/SVM), Τυχαία Δάση (Random Forests), Τυχαία Δέντρα (Extra Trees) και Κλίσης Ανόδου (Gradient Boosting). Με την συνάρτηση `featureAndTrain` εκτελείται αλγόριθμος μηχανικής μάθησης για την κατηγοριοποίηση μέσα από ένα σύνολο δεδομένων. Η συνάρτηση συνοδεύεται με ορισμένα βασικά ορίσματα αναγκαία για την εκτέλεση της όπως το όνομα του φάκελου που περιλαμβάνει τα δεδομένα, το μέγεθος των βραχυπρόθεσμων και ενδιάμεσων πλαισίων, το βήμα που ακολουθούν καθώς και τον τύπο της κατηγοριοποίησης.

Ένας εξίσου σημαντικός παράγοντας στον τομέα της ανάλυσης του ήχου είναι η Παλινδρόμηση, η οποία όπως και η Ομαδοποίηση διαχωρίζει το δείγμα βάσει της διαφορετικότητας των χαρακτηριστικών του. Χαρακτηριστικό παράδειγμα είναι η αναγνώριση συναισθημάτων μέσα από την ομιλία, όπου η συναισθηματική κατάσταση δεν είναι διακριτή κλάση αλλά μια πραγματική τιμή. Αντίστοιχα με την συνάρτηση `featureAndTrain` λειτουργεί η συνάρτηση `featureAndTrainRegression` για την παλινδρόμηση χρησιμοποιώντας τα ίδια βασικά ορίσματα. Η βιβλιοθήκη υποστηρίζει Παλινδρόμησης Διανυσμάτων Υποστήριξης για την εκπαίδευση του δείγματος, με σκοπό την χαρτογράφηση των ηχητικών χαρακτηριστικών σε μια ή περισσότερες μεταβλητές. Όπως και στην Ομαδοποίηση πρέπει να υπάρχει ένας φάκελος, όπου περιέχει το κάθε ηχητικό δείγμα ως ξεχωριστό τμήμα, του οποίου η θέση μπαίνει ως όρισμα στην συνάρτηση. Ο φάκελος επίσης πρέπει να περιλαμβάνει και το αντίστοιχο αρχείο `.csv` (comma-separated values), το οποίο περιέχει τις τιμές των μονάδων μέτρησης για κάθε ηχητικό δείγμα. Η εκπαίδευση με Παλινδρόμηση περιέχει παράμετρο διαδικασίας συντονισμού (tuning procedure), όπου υλοποιείται εκτίμηση διασταυρωμένης επικύρωσης (cross-validation) η οποία εκτελείται μέσω της συνάρτησης `evaluateRegression`. Ωστόσο το μέτρο της απόδοσης μεγιστοποιείται με την χρήση του Μέσου Τετραγωνικού Σφάλματος (Mean Square Error/MSE) που υλοποιεί η συνάρτηση της εκπαίδευσης με Παλινδρόμηση. Παρακάτω δίνεται ο μαθηματικός τύπος Μέσου Τετραγωνικού Σφάλματος:

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

Όπου \hat{Y} είναι πίνακας που αποτελείται από n προβλέψεις και προέρχεται από δείγμα n σημείων δεδομένων για κάθε μεταβλητή. Ο Y_i είναι πίνακας που αποτελείται από τις παρακολουθούμενες τιμές των μεταβλητών που έχουν προβλεφθεί. Επιπρόσθετα στον υπολογισμό του MSE για κάθε τιμή παραμέτρου, η οποία έχει ελεγχθεί, το MSE του εκπαιδευόμενου δεδομένου υπολογίζεται για να παροχή μια μέτρηση μιας σφαλματογόνους υπεραρμογής (overfitting). Τέλος η διαδικασία συντονισμού επιστρέφει το MSE της εκτίμησης του μέσου όρου (average estimator). Είναι μια μέθοδος η οποία πάντα επιστρέφει την μέση τιμή της εκτιμώμενης παραμέτρου, βάσει της εκπαίδευσης των δεδομένων, με σκοπό να καθορίσει ένα σημείο αναφοράς για την απόδοση της μέτρησης.

6.1.6 Gensim

Το Gensim στοχεύει στην επεξεργασία ακατέργαστου και αδόμητου ψηφιακού κειμένου (απλό κείμενο/plain text). Είναι βιβλιοθήκη της Python που προσφέρει μοντελοποίηση θεμάτων (topic modelling), ευρετήριο εγγράφων (Document Indexing) και ανάκτηση πληροφορίας με βάση την ομοιότητα (Similarity Retrieval) για μεγάλο όγκου πληροφορίας κείμενα. Οι αλγόριθμοι που χρησιμοποιεί, όπως πχ η Λανθάνουσα Σηματολογική Ανάλυση (Latent Semantic Analysis/LSA), ανακαλύπτουν την σηματολογική δομή των αρχείων εξετάζοντας τα στατιστικά των λέξεων σε συνένφανιση με ακολουθίες που ανήκουν σε ένα σώμα κειμένων (corpus) προς εκπαίδευση. Είναι μη επιβλεπόμενοι αλγόριθμοι, κάτι που σημαίνει ότι δεν είναι απαραίτητη καμία είσοδο από το χρήστη και το μόνο που χρειάζεται είναι ένα σύνολο αρχείων, απλού κειμένου. Αφού βρεθούν οι ακολουθίες οποιοδήποτε αρχείο απλού κειμένου μπορεί να εκφραστεί ως ένα καινούργιο, σηματολογικής αναπαράστασης και να εξεταστεί για ομοιότητες έναντι άλλων αρχείων.

Αρχικά πρέπει να δημιουργήσουμε το λεξικό και το σώμα κειμένων, το οποίο επιτυγχάνουμε με την χρήση των συναρτήσεων Dictionary και MmCorpus.serialize αντίστοιχα. Η πρώτη συνάρτηση παίρνει ως όρισμα μια λίστα, η οποία αποτελείται από το σύνολο των λέξεων ή ακόμα και συμβόλων που θέλουμε να περιλαμβάνει το λεξικό (προαιρετικά ως δεύτερο όρισμα δέχεται το όριο που περιορίζει τον αριθμό των επιτρεπτών στοιχείων). Σκοπός είναι η χαρτογράφηση των λέξεων με τις διάφορες ταυτότητες που τους αναθέτει η ίδια η συνάρτηση. Όσον αφορά την συνάρτηση για την δημιουργία του σώματος κειμένων, αυτή δέχεται αρκετά ορίσματα όπως το μονοπάτι που θα αποθηκευτεί το ευρετήριο, τη δυνατότητα χαρτογράφησης των ταυτοτήτων σε

λέξεις κλπ. Δύο είναι όμως τα απαραίτητα ορίσματα για να λειτουργήσει η συνάρτηση, το μονοπάτι για το εξαγόμενο αρχείο που θα δημιουργηθεί και το ακατέργαστο σώμα κειμένου σε “Σάκους Λέξεων” μορφότυπο (Bag-of-Words/BoW). Το αποτέλεσμα της διαδικασίας αυτής μπορεί να χρησιμοποιηθεί ως εξαχθέν χαρακτηριστικό του αρχικού κειμένων.

Για να έχουμε μεγαλύτερη επιτυχία από τα εξαγόμενα αυτά χαρακτηριστικά μπορούμε να χρησιμοποιήσουμε μεθόδους όπως το μοντέλο της Συχνότητας του Όρου στο Κείμενο και Άνισης Κατανομής του Όρου στο Κείμενο (Term Frequency-Inverse Document Frequency/ TFIDF) και το μοντέλο Λανθάνουσας Σημασιολογικής Δεικτοδότησης (Latent Semantic Indexing/LSI). Το πρώτο μοντέλο υλοποιείται μέσω της συνάρτησης `TfidfModel` και ουσιαστικά πολλαπλασιάζει μια τοπική συνιστώσα (Συχνότητας του όρου στο Κείμενο) με μια καθολική συνιστώσα (Άνισης Κατανομής του Όρου στο Κείμενο) και ομαλοποιεί τα αποτελέσματα στην μονάδα του μήκους. Ο μαθηματικός τύπος για μη ομαλοποιημένο βάρος όρου i σε ένα αρχείο j ενός σώματος κειμένου από D αρχεία:

$$weight_{i,j} = frequency_{i,j} \times \log_2 \frac{D}{document_freq_i}$$

ή στην γενική του μορφή

$$weight_{i,j} = wlocal(frequency_{i,j}) \times wglobal(document_freq_i, D)$$

Έχει παραμέτρους όπως σώμα κειμένων, ομαλοποίηση, προσδιορισμό συγκεκριμένου λεξικού κλπ., οι οποίες είναι προαιρετικές καθώς και μόνο με το σώμα κειμένου η συνάρτηση μπορεί να υλοποιηθεί. Αντίστοιχα η Λανθάνουσα Σημασιολογική Δεικτοδότηση (Latent Semantic Indexing/LSI) χρησιμοποιείται για την αποσύνθεση ενός μεγάλου σώματος κειμένων με χρήση στατιστικών στοιχείων. Επιλύει δύο από τα σημαντικότερα προβλήματα των ερωτημάτων με λέξεις κλειδιά του Μπουλ (Boolean keyword queries), Από την μια πλευρά πολλαπλές λέξεις έχουν πανομοιότυπη σημασία τα όποια ονομάζεται “συνώνυμα” (synonyms) και από την άλλη λέξεις έχουν πολλαπλές ερμηνείες κάτι που ονομάζεται “πολυσημία” (polysemy). Ομοίως με το προηγούμενο μοντέλο, υποχρεωτική παράμετρος για την λειτουργία του είναι το σώμα κειμένων.

Επιπρόσθετες λειτουργίες προσφέρει με παραμέτρους όπως ο αριθμός των αρχείων που χρησιμοποιείται για κάθε εκπαιδευόμενο κομμάτι, επιλογή για κατανεμημένη λειτουργία κτλ.

6.2 Ανάλυση Μεθοδολογίας

Για να μπορέσουμε να τρέξουμε και να αναπαράγουμε αποτελέσματα με το pyMusicEmotionRecognition εκτελέστηκε η ακολουθία των παρακάτω βημάτων:

- Βήμα 1.** Λήψη και συλλογή δειγμάτων μουσικών δεδομένων. Απομόνωση της πληροφορίας για κάθε μορφότυπο ξεχωριστά.
- Βήμα 2.** Εξαγωγή χαρακτηριστικών και εκπαίδευση για κάθε δείγμα που διαχωρίσαμε σύμφωνα με το Βήμα 2.
 - 2.1 Εξαγωγή χαρακτηριστικών και εκπαίδευση δειγμάτων μουσικού σήματος.
 - 2.2 Εξαγωγή χαρακτηριστικών και εκπαίδευση δειγμάτων μουσικής στιχουργίας.
- Βήμα 3.** Συλλογή αποτελεσμάτων που διακρίνονται ως τα καλύτερα, ξεχωριστά για κάθε μορφότυπο.
- Βήμα 4.** Συλλογή και σύγκριση των αποτελεσμάτων για το κάθε μορφότυπο και εξαγωγή συμπερασμάτων.

6.2.1 Λήψη και Συλλογή Δειγμάτων Μουσικών Δεδομένων

Η λήψη και συλλογή των δειγμάτων διαχωρίζεται σε ηχητικό σήμα σε μορφότυπο mp3 και στίχους (μορφή κειμένου). Πιο συγκεκριμένα από το αρχείο get_YSL γίνεται η συλλογή των στίχων με την συνάρτηση getLyrics, της PyLyrics, και με παραμέτρους το όνομα του καλλιτέχνη και τον τίτλο του κομματιού. Το όνομα του καλλιτέχνη και ο τίτλος του κομματιού συλλέγεται από το spotify συλλέγοντας ταυτόχρονα και τις μονάδες μέτρησης που αντιπροσωπεύουν την έκφραση των συναισθημάτων. Από τις μονάδες που διαθέτει χρησιμοποιήθηκαν: το σθένος, η χορευτικότητα και η ενέργεια (valance, danceability και energy), η επιλογή δεν έγινε τυχαία. Η εύρεση του ζητημένου κομματιού γίνεται με την συνάρτηση search και βάση του τίτλου του (ως τίτλο αναφερόμαστε στον τίτλο που παρέχετε στο βίντεο από τον ιστότοπο του Youtube αφαιρώντας πληροφορίες

όπως καλλιτέχνες που συμμετέχουν στον κομμάτι και διατηρώντας μόνο τον κύριο ερμηνευτή και την ονομασία του κομματιού που ερμηνεύει (για πιο εύστοχη λήψη δειγμάτων λεπτομέρειες καταγράφονται στο αρχείο README.txt). Αφού επιτευχθεί η συλλογή των στίχων αποθηκεύονται, μαζί με το όνομα του καλλιτέχνη και τον τίτλο του κομματιού (καθώς και τις μονάδες μέτρηση των συναισθημάτων), σε ένα αρχείο μορφότυπου tsv.

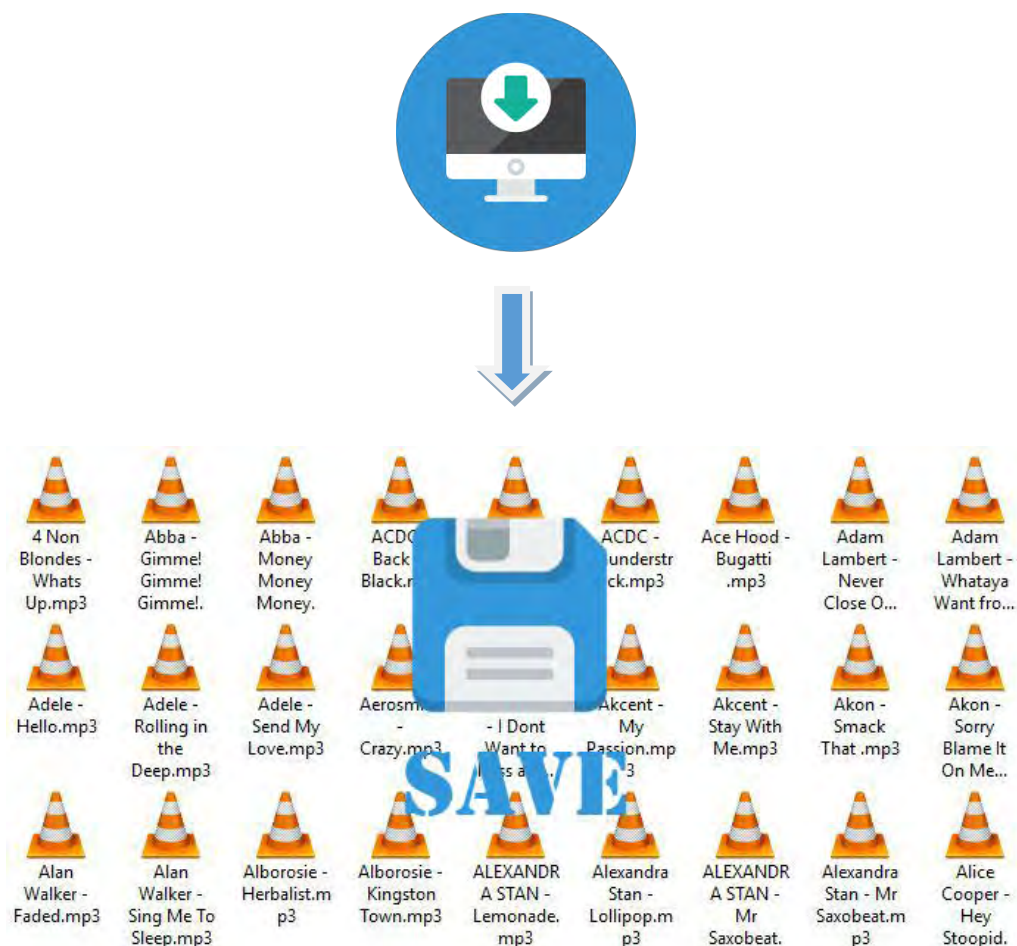
```
Title Artist Track Lyrics Energy Valence Danceability
Yelowolf - Till It's Gone Yelowolf Till It's Gone "I'm not the table you can come and lay your cup down on now I'm not the shoulder for a bag, don't wanna carry your heavy load
Duke Dumont - Ocean Drive Duke Dumont Ocean Drive We're riding down the boulevard We're riding to the dark night, night With half a tank and empty heart Pretending we're in love,
Britney Spears - Womanizer Britney Spears Womanizer Superstar, where you from? How's it going? I know you got a clue what you're doing You can play brand new to all the other chick
Rihanna - Love On The Brain Rihanna Love On The Brain And you got me like oh What you want from me? What you want from me? And I tried to buy your pretty heart, but the price too hi
David Guetta - Dangerous David Guetta Dangerous (feat. Sam Martin) You take me down Spin me around You got me running all the lights Don't make a sound Talk to me now Let me
Mariah Carey - Touch My Body Mariah Carey Touch My Body MC, in the place to be Oh, yeah, oh, yeah, oh, yeah, oh, yeah, oh, yeah, oh, yeah, oh, yeah I know that you been waitin
The Weeknd - Reminder The Weeknd Reminder Recommend play my song on the radio You too busy trying to find that blue-eyed soul I let my black hair grow and my weed smoke And I sw
Muse - Knights Of Cydonia Muse Knights Of Cydonia Come ride with me Through the veins of history I'll show you how God Falls asleep on the job And how can we win When fools can
Three Days Grace - Never Too Late Three Days Grace Never Too Late This world will never be what I expected And if I don't belong, who would've guessed it? I will not leave alone
Metallica - One Metallica One I can't remember anything Can't tell if this is true or dream Deep down inside I feel the scream This terrible silence stops with me Now that the war
Nelly Furtado - Say It Right Nelly Furtado Say It Right In the day, in the night Say it right, say it all You either got it or you don't You either stand or you fall When you
The Black Eyed Peas - Shut Up The Black Eyed Peas Shut Up *Pergie: Shut up Just shut up Shut up, shut up, shut up Shut it up, just shut up Shut up Just shut up Shut up, shut up, shu
The Black Eyed Peas - Hey Mama The Black Eyed Peas Hey Mama A-La la la laa Hey mama, this that shit that make you groove, mama Get on the flow and move your booty mama We the bla
Katy Perry - Dark Horse Katy Perry Dark Horse Juicy J Yeah, y'all know what it is Katy Perry Juicy J Uh huh, let's rave Katy Perry I knew you were You were gonna come to me And
The Weeknd - The Hills The Weeknd The Hills Your man on the road, he doing promo You said keep our business on the low-low I'm just tryna get you out the friend-zone 'Cause you lo
Metallica - Nothing Else Matters Metallica Nothing Else Matters So close, no matter how far Couldn't be much more from the heart Forever trusting who we are And nothing else m
```

Εικόνα 5.Στιγμιότυπου της παράταξης των στοιχείων ενός κομματιού (Τίτλος/Καλλιτέχνης/Όνομα Κομματιού/..) σε μορφή .tsv.

Με παρόμοιο τρόπο λειτουργεί και η λήψη του ηχητικού σήματος με την διαφορά ότι υπάρχει συνθήκη η όποια ελέγχει για την ύπαρξη των μονάδων μέτρησης των συναισθημάτων. Μέσω του αρχείου get_YSL διαβάζει τις διεύθυνσεις (url) των βίντεο, του ιστότοπου YouTube, που είναι αποθηκευμένες στο αρχείο url_file και με την συνάρτηση download και τις ανάλογες παραμέτρους κατεβάζει και αποθηκεύει τα βίντεο στο μορφότυπο mp3.

```
68 ydl = youtube_dl.YoutubeDL(ydl_opts)
69
70 ydl.download([url]) #Download Video Sound with YouTube URL
```





Σχήμα 9. Στο παραπάνω διάγραμμα απεικονίζετε ο κώδικας που εκτελείτε για την λήψη των αρχείων του ήχου καθώς και η λήψη και αποθήκευση σε τοπικό αρχείο του ηλεκτρονικού υπολογιστή.

Τα κομμάτια αποθηκεύονται μεμονωμένα στο φάκελο “/songs” σε αντίθεση με τους στίχους και τις μονάδες μέτρησης των αισθημάτων που βρίσκονται στο ίδιο αρχείο .tsv. Η αμέσως επομένη διαδικασία διαχωρίζει τους στίχους τους κάθε κομματιού μεμονωμένα σε αρχεία txt στο φάκελο “/lyrics”. Όμοια αποθηκεύονται σε αρχεία μορφότυπου .csv, οι τρεις μονάδες που αντιπροσωπεύουν την έκφραση των συναισθημάτων: σθένος, χορευτικότητα και ενέργεια. Αυτά τα τρία .csv αρχεία αποθηκεύονται εις διπλούν μια φορά για το φάκελο “/songs” και μια για τον φάκελο “/lyrics”.

6.2.2 Εξαγωγή Χαρακτηριστικών και Εκπαίδευση

Για κάθε ένα από κομμάτια τόσο όσον αφορά τον ήχο όσο και τους στίχους υπάρχει η αντίστοιχη συνάρτηση `dirWavFeatureExtraction` και `dirLyrFeatureExtraction` όπου

εξάγουν τα χαρακτηριστικά και αρχειοθετούνται σε μια λίστα. Κάθε λίστα, μια για το ηχητικό σήμα και μια για το κείμενο, αντιστοιχεί σε ένα μουσικό κομμάτι. Η υλοποίηση της κάθε μιας διαφέρει αλλά η μορφή των χαρακτηριστικών που παραδίδεται για εκπαίδευση είναι ίδια. Κάθε χαρακτηριστικό έχει μια πραγματική τιμή. Στην περίπτωση του ήχου αντιπροσωπεύει κάποιο στιγμιότυπο του ηχητικού κομματιού, το οποίο προέκυψε από την διαδικασία της Βραχυπρόθεσμης και Ενδιάμεση Εξαγωγής. Αντίστοιχα στην περίπτωση των στίχων κάποια λέξη από τους στίχους ενός κομματιού, η οποία προέκυψε από την διαδικασία της στατιστικής ανάλυση για την δημιουργία σώματος κειμένου και των μοντέλων που εφαρμοστήκαν σε αυτό για την αύξηση της απόδοσης του.

Όπως αναφέρθηκε στην προηγούμενη παράγραφο, τα αποτελέσματα της εξαγωγής καταλήγουν στην ίδια μορφή, που είναι ένα σύνολο από λίστες. Και για τις δύο περιπτώσεις ο τρόπος εκπαίδευσης ήταν ο ίδιος και η υλοποίηση έγινε μέσω της συνάρτησης `featureAndTrainRegression`, η οποία αφού διαβάσει τα δεδομένα καλεί την αντίστοιχη συνάρτηση για εξαγωγή των χαρακτηριστικών (`dirWavFeatureExtraction` και `dirLyrFeatureExtraction` για ήχο και τους στίχους αντίστοιχα) και στα αποτελέσματα που λαμβάνει εφαρμόζει τις τεχνικές μηχανικής μάθησης και τον υπολογισμό του σφάλματος. Η εκτέλεση αποδίδει ξεχωριστά αποτελέσματα για κάθε μια από τις μονάδες Σθένους, Ενέργειας και Χωρητικότητας. Επιπροσθέτως η συνάρτηση `featureAndTrainRegression` εκτός από την υλοποίηση της εξαγωγής χαρακτηριστικών, για κάθε ένα από τα μορφότυπα, διαφέρει και στα ορίσματα. Συγκεκριμένα στην περίπτωση των στίχων δεν παίρνει ως όρισμα το μέγεθος των παραθύρων κάτι που κάνει για τον ήχο, (αναγκαίο στη διαδικασία της εξαγωγής) ενώ ισχύει το αντίθετο για τον θεματικό αριθμό του LSI μοντέλου.

Τέλος πρέπει να διευκρινίσουμε ότι οι διαδικασίες εκπαίδευσης και εξαγωγής εκτελούνται από διαφορετικά αρχεία, για τον ήχο μέσω του `train_mp3` ενώ για τους στίχους μέσω του `train_lyrics`, όπου στο καθένα αντιστοιχούν και οι ανάλογες από τις παραπάνω συναρτήσεις. Το τελικό αποτέλεσμα που βλέπουμε από την εκπαίδευση είναι το MSE, όσο μικρότερο είναι το σφάλμα τόσο μεγαλύτερη επιτυχία έχει η εκπαίδευση.

6.2.3 Συλλογή και Σύγκριση Αποτελεσμάτων

Τελευταίο χρησιμοποιείται το αρχείο `get_train_results`, με το οποίο συλλέγουμε τα καλύτερα αποτελέσματα που προέκυψαν από τις διάφορες εκτελέσεις που κάναμε για την εκπαίδευση του δείγματος. Είναι προαιρετική διαδικασία η οποία διευκολύνει τον χρήστη να διακρίνει τα καλύτερα αποτελέσματα ανάμεσα στο πλήθος των συνολικών αποτελεσμάτων. Διαχωρίζει καθένα από τα χαρακτηριστικά ξεχωριστά και παίρνει ως όρισμα το μονοπάτι του φακέλου του μορφότυπο που επιθυμούμε (π.χ. για τον ήχο “SongsResults”). Κατά την διάρκεια της εκπαίδευσης τα αποτελέσματα εκτυπώνονται στο Τερματικό (Terminal) του Λειτουργικού Συστήματος (Operating System/OS) του λειτουργικού και σε αρχείο, με όνομα `FeatureAndTrainRegressionResults.txt` μορφότυπου, το MSE αποτέλεσμα της Παλινδρόμησης. Από το `FeatureAndTrainRegressionResults`, με την συνάρτηση `getResults`, συλλέγονται οι καλύτερες τιμές από κάθε εκπαίδευση. Συνεπώς μπορούμε να συγκρίνουμε τα αποτελέσματα μέσα από τα `.csv` αρχεία που δημιουργούνται από το `get_train_results`.

6.3 Αποτελέσματα

Μέσα από την εκτέλεση του προγράμματος λάβαμε τα εξής καλύτερα αποτελέσματα που αναγράφονται στους ακόλουθους πίνακες. Οι πρώτοι τρεις πίνακες είναι τα αποτελέσματα από την εκπαίδευσης του ηχητικού σήματος. Στις πρώτες τέσσερις στήλες εμφανίζονται τα μεγέθη των Βραχυπρόθεσμο και Ενδιάμεσων παράθυρων καθώς και τα αντίστοιχα βήματα που ακολουθούν. Στην πέμπτη στήλη αναφέρεται η παράμετρος της Εκτίμησης Διασταυρωμένης Επικύρωσης. Τέλος η τελευταία στήλη αντιπροσωπεύει το Μέσω Τετραγωνικό Σφάλμα, το οποίο μας υποδεικνύει την αποτελεσματικότητα της εκπαίδευσης.

Πίνακας 2.0 πίνακας με τις παραμέτρους και το αποτέλεσμα του Μέσου Τετραγωνικού Σφάλματος από την εκπαίδευση του ηχητικού σήματος για το Σθένος.

Σθένος					
Ενδιάμεσο Μέγεθος Παραθύρου	Ενδιάμεσο Βήμα	Βραχυπρόθεσμο Μέγεθος Παραθύρου	Βραχυπρόθεσμο Βήμα	Παράμετρος Διασταυρωμένης Επικύρωσης	Μέσο Τετραγωνικό Σφάλμα
0.5	0.25	0.02	0.01	0.1	0.0322

Σθένος					
Ενδιάμεσο Μέγεθος Παραθύρου	Ενδιάμεσο Βήμα	Βραχυπρόθεσμο Μέγεθος Παραθύρου	Βραχυπρόθεσμο Βήμα	Παράμετρος Διασταυρωμένης Επικύρωσης	Μέσο Τετραγωνικό Σφάλμα
0.5	0.25	0.04	0.02	0.005	0.0308
0.5	0.25	0.05	0.025	0.1	0.0308
0.5	0.25	0.1	0.05	0.01	0.0306
0.5	0.25	0.2	0.1	0.01	0.0318
2.0	1.0	0.02	0.01	0.1	0.0357
2.0	1.0	0.04	0.02	0.005	0.0334
2.0	1.0	0.05	0.025	0.01	0.0344
2.0	1.0	0.1	0.05	0.001	0.035
2.0	1.0	0.2	0.1	0.001	0.0355
5.0	2.5	0.02	0.01	0.01	0.0382
5.0	2.5	0.04	0.02	0.005	0.0352
5.0	2.5	0.05	0.025	0.005	0.0364
5.0	2.5	0.1	0.05	0.001	0.0371
5.0	2.5	0.2	0.1	0.001	0.0364
10.0	5.0	0.02	0.01	0.01	0.0401
10.0	5.0	0.04	0.02	0.01	0.0376
10.0	5.0	0.05	0.025	0.005	0.0381
10.0	5.0	0.1	0.05	0.001	0.0378
10.0	5.0	0.2	0.1	0.001	0.0376

Πίνακας 3. Ο πίνακας με τις παραμέτρους και το αποτέλεσμα του Μέσου Τετραγωνικού Σφάλματος από την εκπαίδευση του ηχητικού σήματος για την Χορευτικότητα.

Χορευτικότητα					
Ενδιάμεσο Μέγεθος Παραθύρου	Ενδιάμεσο Βήμα	Βραχυπρόθεσμο Μέγεθος Παραθύρου	Βραχυπρόθεσμο Βήμα	Παράμετρος Διασταυρωμένης Επικύρωσης	Μέσο Τετραγωνικό Σφάλμα
0.5	0.25	0.02	0.01	0.01	0.0092
0.5	0.25	0.04	0.02	0.01	0.009
0.5	0.25	0.05	0.025	0.001	0.0093
0.5	0.25	0.1	0.05	0.005	0.0093
0.5	0.25	0.2	0.1	0.005	0.0102
2.0	1.0	0.02	0.01	0.01	0.0101
2.0	1.0	0.04	0.02	0.001	0.0096

Χορευτικότητα					
Ενδιάμεσο Μέγεθος Παραθύρου	Ενδιάμεσο Βήμα	Βραχυπρόθεσμο Μέγεθος Παραθύρου	Βραχυπρόθεσμο Βήμα	Παράμετρος Διασταυρωμένης Επικύρωσης	Μέσο Τετραγωνικό Σφάλμα
2.0	1.0	0.05	0.025	0.005	0.0101
2.0	1.0	0.1	0.05	0.001	0.0102
2.0	1.0	0.2	0.1	0.005	0.0106
5.0	2.5	0.02	0.01	0.005	0.0105
5.0	2.5	0.04	0.02	0.01	0.0103
5.0	2.5	0.05	0.025	0.001	0.0102
5.0	2.5	0.1	0.05	0.001	0.0106
5.0	2.5	0.2	0.1	0.01	0.011
10.0	5.0	0.02	0.01	0.005	0.0112
10.0	5.0	0.04	0.02	0.01	0.0108
10.0	5.0	0.05	0.025	0.001	0.0108
10.0	5.0	0.1	0.05	0.001	0.011
10.0	5.0	0.2	0.1	0.005	0.0109

Πίνακας 4. Ο πίνακας με τις παραμέτρους και το αποτέλεσμα του Μέσου Τετραγωνικού Σφάλματος από την εκπαίδευση του ηχητικού σήματος για την Ενέργεια.

Ενέργεια					
Ενδιάμεσο Μέγεθος Παραθύρου	Ενδιάμεσο Βήμα	Βραχυπρόθεσμο Μέγεθος Παραθύρου	Βραχυπρόθεσμο Βήμα	Παράμετρος Διασταυρωμένης Επικύρωσης	Μέσο Τετραγωνικό Σφάλμα
0.5	0.25	0.02	0.01	10.0	0.0135
0.5	0.25	0.04	0.02	0.005	0.0132
0.5	0.25	0.05	0.025	0.01	0.0137
0.5	0.25	0.1	0.05	0.01	0.0132
0.5	0.25	0.2	0.1	0.1	0.0124
2.0	1.0	0.02	0.01	0.01	0.0147
2.0	1.0	0.04	0.02	0.005	0.0135
2.0	1.0	0.05	0.025	0.01	0.0138
2.0	1.0	0.1	0.05	0.005	0.0132
2.0	1.0	0.2	0.1	0.05	0.0129
5.0	2.5	0.02	0.01	0.01	0.0151
5.0	2.5	0.04	0.02	0.05	0.0137
5.0	2.5	0.05	0.025	0.01	0.0136

Ενέργεια					
Ενδιάμεσο Μέγεθος Παραθύρου	Ενδιάμεσο Βήμα	Βραχυπρόθεσμο Μέγεθος Παραθύρου	Βραχυπρόθεσμο Βήμα	Παράμετρος Διασταυρωμένης Επικύρωσης	Μέσο Τετραγωνικό Σφάλμα
5.0	2.5	0.1	0.05	0.01	0.0132
5.0	2.5	0.2	0.1	0.01	0.0128
10.0	5.0	0.02	0.01	0.005	0.0151
10.0	5.0	0.04	0.02	0.05	0.014
10.0	5.0	0.05	0.025	0.001	0.0139
10.0	5.0	0.1	0.05	0.001	0.0132
10.0	5.0	0.2	0.1	0.005	0.0127

Στους τρεις αμέσως επομένους πίνακες αναγράφονται τα αποτελέσματα της εκπαίδευσης με τους στίχους των κομματιών. Η πρώτη στήλη περιέχει το θεματικό αριθμό του LSI μοντέλου. Οι δυο τελευταίες στήλες περιέχουν την παράμετρος της Εκτίμησης Διασταυρωμένης Επικύρωσης και το Μέσω Τετραγωνικό Σφάλμα αντίστοιχα με τους πίνακες των αποτελεσμάτων του ηχητικού σήματος.

Πίνακας 5.0 πίνακας με της παραμέτρους και το αποτέλεσμα του Μέσου Τετραγωνικού Σφάλματος από την εκπαίδευση των στίχων για το Σθένος.

Σθένος		
Θεματικός Αριθμός LSI	Παράμετρος Διασταυρωμένης Επικύρωσης	Μέσο Τετραγωνικό Σφάλμα
25.0	10.0	0.0493
50.0	0.005	0.0504
100.0	0.001	0.0505
1000.0	0.25	0.0548

Πίνακας 6.0 πίνακας με της παραμέτρους και το αποτέλεσμα του Μέσου Τετραγωνικού Σφάλματος από την εκπαίδευση των στίχων για την Χορευτικότητα.

Χορευτικότητα		
Θεματικός Αριθμός LSI	Παράμετρος Διασταυρωμένης Επικύρωσης	Μέσο Τετραγωνικό Σφάλμα
25.0	0.5	0.0177
50.0	0.001	0.0187
100.0	0.001	0.019
1000.0	0.05	0.022

Πίνακας 7.0 πίνακας με της παραμέτρους και το αποτέλεσμα του Μέσου Τετραγωνικού Σφάλματος από την εκπαίδευση των στίχων για την Ενέργεια.

Ενέργεια		
Θεματικός Αριθμός LSI	Παράμετρος Διασταυρωμένης Επικύρωσης	Μέσο Τετραγωνικό Σφάλμα
25.0	0.005	0.0291
50.0	0.1	0.0286
100.0	0.001	0.0311
1000.0	0.1	0.0293

ΣΥΜΠΕΡΑΣΜΑΤΑ

Όπως αναφέρθηκε αρχικά, στόχος της Πτυχιακής ήταν η μελέτη μέσα από την εκπαίδευση διαφορετικών μορφοτύπων, του κειμένου (στίχοι) και του ηχητικού σήματος. Μέσω της σύγκρισης των διαφορετικών αποτελεσμάτων αντιλαμβανόμαστε την ποιότητα και το ποσοστό πληροφορίας που μας παρέχει κάθε ένα από τα μορφότυπα που χρησιμοποιήθηκαν. Η σύγκριση έγινε μεταξύ τριών μεταβλητών μέτρησης συναισθημάτων, Σθένος, Χορευτικότητα και Ενέργεια. Συγκεκριμένα τα αποτελέσματα κάθε μεταβλητής συγκρίθηκαν με τα αντίστοιχα της μεταβλητής του άλλου μορφότυπου. Για παράδειγμα, τα αποτελέσματα της εκπαίδευσης του ηχητικού σήματος για το σθένος με τα αντίστοιχα αποτέλεσμα του κειμένου κλπ.

Κατά την σύγκριση διαπιστώθηκε ότι το σφάλμα που προέκυψε από την εκπαίδευση του ηχητικού σήματος είναι μικρότερο από αυτό του κειμένου. Από τον πίνακα για παράδειγμα της ενέργειας, βλέπουμε ότι το μικρότερο σφάλμα (συνεπώς και το καλύτερο) που λάβαμε από την εκπαίδευση των στίχων έχει τιμή 0.0124. Η αντίστοιχη βέλτιστη τιμή που λαμβάνουμε από την εκπαίδευση των στίχων είναι 0.0286. Παρατηρούμε ότι η διαφορά βρίσκεται στο δεύτερο δεκαδικό. Η μικρότερη διαφορά μεταξύ των δυο μορφοτύπων εμφανίζεται στη Χορευτικότητα, με την απόσταση μεταξύ τους να μειώνεται στο τρίτο δεκαδικό. Παρόλο που δεν είναι μεγάλη η απόσταση των τιμών τους η διαφορά είναι υπαρκτή και επιδέχεται βελτίωση.

Σε ένα ποσοστό, αυτή η διαφορά των αποτελεσμάτων των δυο μορφοτύπων ήταν αναμενόμενη καθώς η θεματολογία μεταξύ των διαφορετικών ειδών μουσικής είναι μεγάλη, δίχως να συμβαδίζει απαραίτητα με τη μουσική. Για παράδειγμα, σε ένα χιπ χοπ κομμάτι το περιεχόμενό αναφέρεται στις δυσκολίες της ζωής ενώ ένα άλλο να περιγράφει χρήματα, δύναμη κλπ. Επιπρόσθετα ο τρόπος (τόνος, ταχύτητα κλπ.) έκφρασης οποιαδήποτε λέξη δίνει διαφορετική σημασία στην ερμηνεία της (πχ η αύξηση της ένταση της φωνής, στην πλειοψηφία της, δηλώνει οργή), πληροφορία την αποκρύπτει το κείμενο από μόνο του.

Τα αποτελέσματα που προέκυψαν αποδόθηκαν βάσει της μεθοδολογίας που περιγράφεται στο κεφάλαιο 6. Είναι δυνατόν να δοθούν αποτελέσματα, με μεγαλύτερο βαθμό επιτυχίας, με τη χρήση άλλων μεθόδων ή με την βελτιστοποίηση της παρούσας μεθόδου. Επίσης οι υποδομές που προσφέρει η παρόν εργασία, παρέχουν εφόδια για την επέκταση του εύρους της έρευνας και στο τομέα της εικόνας και του βίντεο. Είναι ακόμα εφικτό να γίνει η σύγκριση με τον συνδυασμό χαρακτηριστικών διαφορετικού μορφότυπου (πχ χαρακτηριστικά ήχου-κειμένου).

ΠΙΝΑΚΑΣ ΟΡΟΛΟΓΙΑΣ

Ξενόγλωσσος όρος	Ελληνικός Όρος
Accelerando	επιταχυνόμενος
aesthetic emotions	καλαισθητα συναισθήματα
Allegro	ζωηρός
anchor configurations	διαμόρφωση άγκυρας
anchor matches	ταύτιση αγκύρας
Anger	θυμός
Anticipation	προσδοκία
Anxiety	νευρικότητα
Appraisal	αποτίμηση
Arousal	διέγερση
Arpeggio	αρπέτζιο
Arpeggio	αρπέτζιο
associated frequency	σχετική συχνότητα
Attack	Δράση (1 ^η φάση της περιβάλλουσα ενός ήχου)
Audio	ήχος
average estimator	μέσου όρου
Average Zero-Crossing Rate	Ρυθμού Αλλαγής Πρόσημου
Bag-of-Words	“Σάκος Λέξεων”
Bandpass	ζωνοπερατό
Chroma	χρώμα
Chroma Deviation	Χρωματική Απόκλιση
Chroma Vector	Χρωματικός Πίνακας
Classification	Κατηγοριοποίηση
Cognitivist	Μη-Συναισθηματική Ιδεολογία
content based	βάσει το περιεχόμενο
Corpus	σώμα κειμένων
cross-validation	εκτίμηση διασταυρωμένης επικύρωσης
Danceability	χωρητικότητα
DataFrame	Πλαίσιο Δεδομένων
Decay	Μερική Εξασθένιση (3 ^η φάση της περιβάλλουσα ενός ήχου)
deliberateness, intentionality	προδιάθεση, σκοπιμότητα
Dependent	εξάρτηση
Disgust	αποστροφή
document indexing	ευρετήριο εγγράφων
Downsampled	μείωση μεγέθους δείγματος
Duration	διάρκεια
Emotivist	Συναισθηματική Ιδεολογία
Energy	Ενέργεια
Entropy of Energy	Εντροπία της Ενέργειας
equal-tempered scale	νότες ισοσκελής κλίμακας
Expectations	προσδοκίες
Explanatory	επεξηγηματικότητα
Extra Trees	Τυχαία Δέντρα
Extraction	εξαγωγή
Fear	φόβος
Feature	χαρακτηριστικό

Ξενόγλωσσος όρος	Ελληνικός Όρος
Filter	φίλτρο
filter bank	τράπεζα φίλτρων
Format	μορφότυπο
Frame	παράθυρο
grace note	επέρειαση
grace note	επέρειαση
Gradient Boosting	Κλίσης Ανόδου
Hold	Διατήρηση στο Μέγιστο Επίπεδο (2 ^η φάση της περιβάλλουσα ενός ήχου)
Hyperplane	υπερεπίπεδο
Hypothesis	υπόθεση
Imagination	φαντασία
Independent	ανεξαρτησία
Intensity	ένταση
Intensity	ένταση
Joy	χαρά
k Nearest Neighbor	κ Πλησιέστερος Γείτονας
Keyword	λέξη κλειδί
Latent Semantic Indexing	Λανθάνουσα Σημασιολογική Δεικτοδότηση
Mean Square Error	Μέσω Τετραγωνικό Σφάλμα
Mel Frequency Cepstral Coefficients	Συντελεστές Συχνότητας με Σάφματα Μελ
mid-term	ενδιάμεσο (είδος χαρακτηριστικού)
Multirate	πολλαπλών τιμών
Multiscale	πολλαπλής κλίμακας
Music Information Retrieval	Ανακτώσης Μουσικής Πληροφορία
Musical Instrument Digital Interface	Ψηφιακή Διασύνδεση Μουσικών Οργάνων
musical onset time	μουσικός χρόνος έναρξης
Needs	ανάγκες
non-overlapping	μη επικαλυπτόμενα
Notation Software	Λογισμικό Μουσικής Σημειογραφίας
Operating System	Λειτουργικού Συστήματος
Optical Music Recognition	Οπτική Μουσική Αναγνώριση
Overfitting	σφαλματογόνους υπεραρμογής
Overlapping	επικαλυπτόμενα
Pitch	τονικότητα
pitch class	τονική κλάση
Pleasantness	ικανοποίηση
Polysemy	πολυσημία
Prediction	πρόβλεψη
Predictor	προβλέψιμο
pseudo-emotions	ψευδό-συναισθήματα
Pulse Code Modulation	Παλμοκωδική Διαμόρφωση
quality factor	παράγοντα ποιότητας
Queries	ερωτήματα
Random Forests	Τυχαία Δάση
Random Trees	Τυχαία Δέντρα
reaction	ανταπόκριση
Regression	Παλινδρόμηση

Ξενόγλωσσος όρος	Ελληνικός Όρος
Release	Αποδέσμευση
response	απόκριση
ritardando	επιβραδυνόμενος
sadness	θλίψη
saturation	κορεσμός
score	παρτιτούρα
scorewriter	λογισμικό σημειογραφίας
Segmentation	Διαμέριση
Sequencer	Λογισμικό Μουσικής Εγγραφής
series	ακολουθία
short-term	βραχυπρόθεσμο (είδος χαρακτηριστικού)
Similarity Retrieval	Ανάκτηση Πληροφορίας με Βάση την Ομοιότητα
softness	ομαλότητα
solemnity	επισημότητα
sonority	ηχηρότητα
Spectral Centroid	Κεντρικό Φάσμα
Spectral Entropy	Εντροπία Φάσματος
Spectral Flux	Φασματική Διακύμανση
Spectral Rolloff	Φασματική Αναίρεση
Spectral Spread	Διευρυμένο Φάσμα
subband	υποζώνη
Supervised Learning	Επιβλεπόμενη Μάθηση
Support Vector Machines	Μηχανής των Διανυσμάτων Υποστήριξης
Support Vector Regression	Διάνυσμα Υποστήριξης Παλινδρόμησης
surprise	έκπληξη
Sustain	Διατήρηση ενός Μέσου Επιπέδου (4 ^η φάση της περιβάλλουσα ενός ήχου)
synonyms	συνώνυμα
tempo	ρυθμός
tempo	ρυθμός
tension	ένταση
Term Frequency-Inverse Document Frequency	Συχνότητα του Ορού στο Κείμενο και Άνισης Κατανομής του Όρου στο Κείμενο
Terminal	Τερματικό
Test Set	Σύνολο Ελέγχου
text-based	βάσει το κείμενο
theory	θεωρία
timestamp	χρονοσφραγίδα
Tokenization	Ανάθεση Λεκτικής Μονάδας
tone height	τονικό ύψος
tone quality	χροιά ήχου
topic modeling	μοντελοποίηση θεμάτων
Training Set	Σύνολο Εκπαίδευσης
trill	τρίλισμα
triviality	ανεπισημότητα
trust	εμπιστοσύνη
tuning procedure	διαδικασίας συντονισμού
unpleasantness	δυσαρέσκεια
valance	σθένος

Ξενόγλωσσος όρος	Ελληνικός Όρος
variable	μεταβλητή
Visualization	Απεικόνιση
Zero Crossing Rate	Ρυθμός Αλλαγής Προσήμου

ΣΥΝΤΜΗΣΕΙΣ – ΑΡΚΤΙΚΟΛΕΞΑ – ΑΚΡΩΝΥΜΙΑ

csv	comma-separated values
BoW	Bag-of-Words
CENS	Chroma Energy Normalized Statistics
CT	Continuous-Time
dB	decibel
DT	Discrete-Time
Extra Trees	Extremely Randomized Trees
ITPRA	Imagination-Tension-Prediction-Reaction-Appraisal
kNN	k Nearest Neighbor
LSA	Latent Semantic Analysis
LSI	Latent Semantic Indexing
MFCCs	Mel Frequency Cepstral Coefficients
MIDI	Musical Instrument Digital Interface
MIR	Music Information Retrieval
MP3	Moving Picture Experts Group (MPEG) Audio Layer III
MsDTW	Multiscale DTW
MSE	Mean Square Error
OMR	Optical Music Recognition
OS	Operating System
PCM	Pulse Code Modulation
SFM	Standard MIDI File
STMSP	Short-Time Mean-Square Power
SVM	Support Vector Machines
SVR	Support Vector Regression
TF-IDF	Term Frequency-Inverse Document Frequency
tsv	tab-separated values
url	uniform resource locator

ΑΝΑΦΟΡΕΣ

- [1] Bolton, D. (2018). rg3/youtube-dl. [Ιστότοπος] github.com. Διαθέσιμο στο: <https://github.com/rg3/youtube-dl/blob/master/README.md> [Προσπελάστηκε 31 Aug. 2018].
- [2] Lamere, P. (2014). Spotipy. [Ιστότοπος] Spotipy.readthedocs.io. Διαθέσιμο στο: <http://spotipy.readthedocs.io/en/latest/> [Προσπελάστηκε 31 Aug. 2018].
- [3] Bora, P. (2015). geekpradd/PyLyrics. [Ιστότοπος] github.com. Διαθέσιμο στο: <https://github.com/geekpradd/PyLyrics> [Προσπελάστηκε 31 Aug. 2018].
- [4] Lambda Foundry and PyData Development Team (2012). Pandas: Python Data Analysis Library. [Ιστότοπος] pandas.pydata.org. Διαθέσιμο στο: <http://pandas.pydata.org/> [Προσπελάστηκε 31 Aug. 2018].
- [5] Giannakopoulos, T. (2004). pyAudioAnalysis. [Ιστότοπος] github.com. Διαθέσιμο στο: <https://github.com/tyiannak/pyAudioAnalysis/wiki> [Προσπελάστηκε 31 Aug. 2018].
- [6] Balz, A. (1914). Music and Emotion. *The Journal of Philosophy, Psychology and Scientific Methods*, 11(9), p.236.
- [7] Zentner, M., Grandjean, D. and Scherer, K. (2008). Emotions evoked by the sound of music: Characterization, classification, and measurement. *Emotion*, 8(4), pp.494-521.
- [8] Shafron, G. R. (2010). The Science and Psychology Behind Music and Emotion. *Journal of Young Investigators*, 20(5), 1-22.
- [9] Gomes, O. (2017). Plutchik and Economics: Disgust, Fear, and, Oh Yes, Love. *Journal of Young Investigators*, 22(1), 37-63.
- [10] Giannakopoulos, T. (2015). pyAudioAnalysis: An Open-Source Python Library for Audio Signal Analysis. *PLOS ONE*, 10(12), p.e0144610.
- [11] Hewitt, M. (2008). *Music Theory for Computer Musicians*. 1st ed. Boston, MA, United States: Course Technology PTR.
- [12] Müller, M. (2007). *Information Retrieval for Music and Motion*. 2007 ed. Berlin, Germany: Springer-Verlag Berlin and Heidelberg GmbH & Co. KG.
- [13] Giannakopoulos, T. and Pikrakis, A (2014). *Introduction to Audio Analysis: A MATLAB (R) Approach*. 1st ed. London, United Kingdom: ELSEVIER SCIENCE & TECHNOLOGY.
- [14] Trnka, R. and Balcar, K. (2011). *Re-Constructing Emotional Spaces: From Experience to Regulation*. 2011 ed. <https://www.researchgate.net/publication/258378370> : Prague Psychosocial Press.
- [15] Schroeder, L., Sjoquist, D. and Stephan, P. (1986). *Understanding regression analysis*. 1st ed. Beverly Hills, California: Sage Publications, Inc.
- [16] Riess Jones, M., N. Popper, A. and R. Fay, R. (2010). *Music Perception*. 2010 ed. New York, NY, United States: Springer-Verlag New York Inc.